Document Title:	**Evaluation of the Honest Opportunity Probation with Enforcement Demonstration Field Experiment (HOPE DFE), Final Report**
Author(s):	**Pamela K. Lattimore, Debbie Dawes, Doris L. MacKenzie, Gary Zajac**
Document Number:	**251758**
Date Received:	**June 2018**
Award Number:	**2011-RY-BX-0003**

NIJ Grant Number 2011-RY-BX-0003
RTI Project Number 021366

Evaluation of the Honest Opportunity Probation with Enforcement Demonstration Field Experiment (HOPE DFE)

Final Report

February 2018

Prepared for

National Institute of Justice
810 7th Street NW
Washington, DC 20531

Prepared by

Pamela K. Lattimore
Debbie Dawes
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709

Doris L. MacKenzie
Gary Zajac
Justice Center for Research
Pennsylvania State University
University Park, PA 16801

RTI INTERNATIONAL

Acknowledgements

The evaluation of the Honest Opportunity Probation with Enforcement (HOPE) Demonstration Field Experiment (DFE) was funded by grant 2011-RY-BX-0003 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect those of the U.S. Department of Justice.

The evaluation team extends our many thanks to those who contributed to this evaluation. In particular, the team thanks the HOPE demonstration sites' judges and court staff, probation administrators and officers, and other individuals in the DFE sites who were critical to the implementation of the HOPE programs and who assisted the evaluation team throughout the study period. The evaluation team would also like to extend many thanks to Ed Banks (BJA) and Eric Martin (NIJ) for their support; and the HOPE technical assistance providers Angela Hawken, Jonathan Kulick, and their team (formerly at Pepperdine University and now at New York University).

Principal Investigators
Pamela K. Lattimore, RTI International
Doris L. MacKenzie, Pennsylvania State University

Report Contributors
Elaine Arsenault, Pennsylvania State University
Alan Barnosky, RTI International
Susan Brumbaugh, RTI International
Joel Cartwright, RTI International
Alexander Cowell, RTI International
Debbie Dawes, RTI International
Pamela K. Lattimore, RTI International
Doris L. MacKenzie, Pennsylvania State University
Derek Ramirez, RTI International
Stephen Tueller, RTI International
Gary Zajac, Pennsylvania State University

Contents

Exhibits

Abstract

Purpose: The multi-site evaluation of the Honest Opportunity Probation with Enforcement Demonstration Field Experiment (HOPE DFE) was a four-site, randomized controlled trial replicating a Hawaii probation program widely touted as successful in reducing drug use, violations, and reincarceration. HOPE is based on "swift, certain, and fair" principles—beginning with a warning hearing from a judge and requiring strict adherance to supervision requirements, including random drug testing, with all violations followed by hearings and jail sanctions; treatment is for those who repeatedly fail random tests. Grants and technical assistance were provided to the sites (Saline County, Arkansas; Essex County, Massachusetts; Clackamas County, Oregon; Tarrant County, Texas) by the Bureau of Justice Assistance to facilitate implementation. The evaluation documented implementation and fidelity; tested outcomes, primarily recidivism; and estimated costs.

Research Subjects: 1,504 HOPE-eligible individuals were randomly assigned to HOPE or to probation as usual (PAU) between August 2012 and September 2014. Most were male (81%), white (69%), and high risk (55%). On average, they were 31 years at study enrollment, with 7 prior arrests and 3.5 prior convictions. Subject characteristics varied across the sites. For example, study participants were younger at first arrest in Texas than Massachusetts (19 versus 27 years) and had more prior convictions in Massachusetts than in Arkansas and Texas (6 versus about 2).

Methods: The evaluation team established procedures with each site for identifying HOPE-eligible probationers and implementing random assignment. Data collection included site visits and document review for the process evaluation, as well as analysis of fidelity data. For the outcome and cost evaluation, administrative data were collected from local and state agencies and three waves of interviews were conducted with study participants. Oral swab drug tests were administered during the second and third interviews for individuals in the community and who consented. A substudy was conducted that enlisted randomly selected subjects in a telephone component that asked subjects to call in weekly and answer a short set of questions to assess whether attitudinal changes occurred over the course of HOPE participation.

Results: Implementation fidelity was good to excellent in the DFE sites, showing adherance to guidelines for warning and violation hearings, random drug testing, and responses to violations. Of the eleven metrics measured, the sites had the greatest difficulty bringing a violator to a violation hearing within 3 days of the violation, although three-quarters did have a hearing within 1 week. Overall, there was strong buy-in to the HOPE concept and implementation was facilitated by existing agency cooperation, prior experience with HOPE-like programs, and organizational linkages between probation and the court. Challenges in some sites included resource constraints—even with grant funding—and conflict with existing probation culture.

HOPE probationers were more likely to have a violation and had more violations than PAU probationers, including more than twice as many drug-related violations accompanying the more than five-fold increase in drug testing for HOPE versus PAU probationers. HOPE probationers were less likely to miss a probation officer visit, to fail to pay their fees and fines , and to be violated for a new charge ; but were more likely to have a violation for failing to appear for court . Most sanctions for HOPE probationers were jail days; HOPE probationers were more likely to go to jail , to go more often), and to serve more days total than PAU probationers.

The HOPE model included treatment referral after repeated failed tests and HOPE participants were three times more likely to go to residential treatment . HOPE probationers were also referred to treatment more quickly (overall and in three sites). Drug tests conducted in conjunction with follow-up interviews showed fewer positives for HOPE than PAU probationers.

Recidivism outcomes were similar for the HOPE and PAU groups: 40% of HOPE versus 44% of PAU had a new arrest; 25% of HOPE versus 22% of PAU had a revocation; 49% of HOPE versus 50% of PAU had an arrest or revocation; and 28% of HOPE versus 26% of PAU had a new conviction. There was some variation in rates across sites, but the general conclusions of no differences hold with two exceptions: (1) HOPE probationers were more likely to be revoked in two sites (PAU revocation rates in those sites were about 10%.); and (2) HOPE probationers were more likely to have a new conviction in one site. Lognormal survival models of time to recidivism events confirm the bivariate findings, but revealed one additional finding—HOPE probationers had longer times to revocation in one site.

Cost analyses estimated costs of intake, warning hearings, staffing meetings, office visits, drug tests, violation hearnings, arrests, state and county corrections, and residential treatment. Six-month median costs were significantly higher for HOPE than PAU overall and in four sites and mean costs were higher overall and in three sites. Twelve-month median and mean costs were significantly higher overall and in three sites. Twenty-four-month median and mean costs were significantly higher overall and in one site. Cost differences were driven by treatment and incarceration costs.

Conclusions: Four sites that differed in organizational structures and populations successfully implemented HOPE programs—holding probationers accountable to their conditions of supervision and reducing drug use. Overall, HOPE did not reduce recidivism, as measured by arrest, revocation, and new conviction. More jail days, more residential treatment, and similar (or higher) recidivism resulted in higher (although not always significantly higher) costs for HOPE compared with PAU.

PAU context is important as sites consider whether to implement HOPE or similar programs based on “swift, certain, and fair” principles. PAU revocation rates were low (9% and 13%) in two sites—suggesting limited ability to reduce revocations and that sites with low PAU revocation rates should consider whether to implement procedures to mitigate any potential increases in revocations that would accompany the increased surveillance of HOPE. In at least two sites, revocation could yield only short prison stays (90 days)—suggesting limited opportunities for “prison bed savings” even if revocations were lower with HOPE and a smaller incentive for individuals to comply. PAU was based at least somewhat on Risk-Needs-Response principles in at least two sites—suggesting an additional consideration with respect to the integration of HOPE with PAU. In addition, in one site, probation could use short jail stays on their authority (and did for PAU cases)—suggesting that a HOPE judge was not necessary to enforce conditions. Thus, the similar outcomes may hinge on the “compared to what” aspect of any evalution—in that findings suggest that HOPE worked as well as but not better than PAU. However, given the consistency of findings across four sites that differed in the administration of PAU, there is little to support a conclusion that HOPE or HOPE-like programs will produce substantial improvements over PAU when implemented widely.

Executive Summary

PROBLEM AND PURPOSE

The multi-site evaluation of the Honest Opportunity Probation with Enforcement Demonstration Field Experiment (HOPE DFE) was a four-site, randomized controlled trial to replicate and evaluate a Hawaii probation program widely touted as successful in reducing drug use, violations, and reincarceration. HOPE is based on "swift, certain, and fair" (SCF) principles—beginning with a warning hearing from a judge and requiring strict adherance to supervision requirements, including random drug testing, with all violations followed by hearings and jail sanctions; treatment is for those who repeatedly fail random tests. The HOPE supervision model was developed under the direction of Judge Steven Alm in 2004 as the Hawaii Opportunity Probation with Enforcement program, building on similar efforts that use certain, but non-severe, graduated sanctions to deter probationers from violating supervision conditions[a]. The HOPE model was in contrast to more traditional approaches to probation—Probation as Usual or PAU—in which multiple violations are tolerated until "a last straw" results in revocation—often to a lengthy prison term.

HOPE DFE

The Bureau of Justice Assistance (BJA) provided implementation grants to Saline County, AR; Essex County, MA; Clackamas County, OR; and Tarrant County, TX. BJA also supported a technical assistance provider to help the sites implement HOPE. The National Institute of Justice funded RTI International and the Center for Justice Research, Pennsylvania State University to conduct the evaluation.

The original studies of the Hawaii HOPE program (Hawken and Kleiman, 2009) found significant improvements for HOPE participants at 1-year follow-up, including 14 percentage point reductions in missed probation appointments; 33 percentage point reductions in positive urine tests; 26 percentage point reductions in new arrests; and 8 percentage point reductions in probation revocations. More recently, Hawken and collegues (2016) reported less dramatic long-term effects with no significant differences in the percentage experiencing a new charge (42% of HOPE versus 47% of PAU), although the average number of new charges was less for HOPE than PAU probationers mostly due to fewer drug charges for the HOPE group. Revocations were reduced for the HOPE group (13% versus 27%) and HOPE probationers were more likely to have their probation terminated early—both of which resulted in criminal justice system savings.

[a] Judge Steven Alm has stated that the HOPE model "is a probationer-centered, collaborative strategy among the judge, probation, defense, prosecution, corrections, law enforcement, and treatment providers to effect positive behavioral change in probationers" (Alm, 2016: p. 1196). In personal conversations in late 2017, Judge Alm said that HOPE was predicated on an assumption of a strong evidence-based rehabilitative approach to probation with the sanctions serving as support for this supervision model. This is not the model of HOPE that was spelled out in the solicitation from the Bureau of Justice Assistance describing the program that was to be implemented by grant recipients (https://www.bja.gov/Funding/11HOPEsol.pdf). It is this sanctions-oriented approach to HOPE that was evaluated by the DFE. Further, as is detailed herein, the four probation offices that implemented the HOPE DFE had a variety of standard probation practices that included risk and needs assessment, referrals to treatment, etc. HOPE was overlaid on these standard practices as would have been the case in Hawaii.

A recent quasi-experimental study of Swift and Certain (SAC) implemented statewide in Washington State showed reductions in prison confienement and duration due to fewer revocations (Hamilton, van Wormer, Kigerl, Campbell, and Posey, 2015; also, see Hamilton, Campbell, van Wormer, Kigerl, and Posey, 2016)—although the pre-post design overlapped with the legalization of marijuana in Washington which may have impacted observed differences. Finally, an experimental evaluation of a deterrence-based program to manage high-risk, substance-using probationers in Delaware found that the program did not reduce substance use or crime for program participants (O'Connell, Brent, and Visher, 2016).

On the promise from the original Hawaii HOPE program, multiple states launched HOPE-like programs (e.g., Petranik, 2011). Hawken and colleagues (2016: 24) note that as of January 2015, HOPE or SCF programs had been implemented in 28 states, one Indian nation, and one Canadian province. The HOPE approach has benefited from a charismatic advocate in Judge Steven Alm, who introduced the program in Hawaii; strong advocacy by the researchers who originally studied the Hawaii program; and support by the National Institute of Justice and the Bureau of Justice Assistance (Duriez, Cullen, and Manchak, 2014). Nonetheless, the evidence base for HOPE remains limited (see, for example, Cullen, Pratt, Turanovic, 2016). The HOPE DFE provided an opportunity to replicate the HOPE program and assess whether a program founded in Hawaii could be generalized to the contiguous United States; most importantly, the DFE provided an opportunity to conduct controlled trial tests of HOPE effectiveness.

EVALUATION GOALS

1. Design & implement a randomization process in each of four DFE sites.

2. Conduct a process evaluation to assess implementation fidelity and identify lessons learned.

3. Conduct a rigorous experimental outcome evaluation to determine the effect of the HOPE model on individual probation outcomes.

4. Conduct a cost evaluation to assess the cost effectiveness of HOPE.

RESEARCH DESIGN

Planning for the evaluation occurred over a 9-month period beginning in late 2011 and involved site visits and discussions with stakeholders in each of the sites to establish HOPE program eligibility criteria and random assignment procedures that were appropriate to each site, as well as to negotiate any compensation that was to be offered to individuals who agreed to participate in the evaluation. During this time, the study team also developed, programmed, and tested baseline and follow-up data collection instruments; designed the T-ACASI study and instrumentation; and developed protocols and semi-structured interviews for the process study. Institutional Review Board review and approval were also obtained during this period and the randomized controlled trial was registered with ClinicalTrials.gov (Number NCT01670708). Discussions were also initiated with agencies in each site in anticipation of negotiating data use agreements for obtaining administrative data in support of the outcome evaluation. Preliminary planning for the cost study also was initiated.

During the planning period, the evaluation team also identified and hired research coordinators in the four sites. The research coordinators were in space provided by the local probation offices in three sites (Massachusetts, Texas, Oregon). Because suitable space was not available in Arkansas, the

evaluation rented an office across the street from the court house. The role of the research coordinators was to introduce the study to each HOPE-eligible probation (prior to random assignment) and solicit his or her participation in the interview portion of the study. The research coordinators administered consent procedures, the audio-computer-assisted-self-interview (ACASI) instruments, and oral swab drug tests during follow-up interviews. The research coordinators also served as local liaisons to the evaluation and assisted with administrative data collection in some of the sites.

The evaluation included process, outcome, and cost components and was designed to address the following research questions:

1. Process
   - 1.1. What was the structural context for the implementation of HOPE in the four sites?
   - 1.2. Was HOPE implemented with fidelity in the four sites?
   - 1.3. What lessons were learned for implementation success, replicability, and sustainability?
   - 1.4. How do intensive drug treatment services offered with the HOPE programs compare with the principles of effective offender intervention?
   - 1.5. What were the communication pathways among HOPE stakeholders and did these vary from site to site?
   - 1.6. How did HOPE probationers view their supervision experiences?
2. Outcome
   - 2.1. Does HOPE participation improve compliance with conditions of supervision and reduce violations?
   - 2.2. Does HOPE participation reduce recidivism, measured by arrest, conviction, and probation revocation?
   - 2.3. What is the impact of HOPE on jail days served and prison days sentenced?
   - 2.4. What is the impact of HOPE on drug use?
   - 2.5. Does HOPE participation change potential mediators including dynamic recidivism risk factors such as employment and housing stability?
   - 2.6. Does HOPE participation change attitudes that are potential mediators, including participants' criminal thinking/attitudes, perceptions of locus of control, and perceptions of the criminal justice system fairness/legitimacy?
3. Cost
   - 3.1. What is the cost of starting and implementing HOPE?
   - 3.2. What are the costs and (any) savings and how are these distributed among the agencies (level of government) participating in HOPE?
   - 3.3. Is HOPE cost effective?

Eligibility determination and random assignment to HOPE or PAU. The study team worked closely with each site to identify who would comprise the local target population; to establish how, when, where, and by whom they would be identified; and to implement an appropriate point and method for randomizing HOPE-eligible probationers to either HOPE or PAU. Eligibility criteria included:

- Risk [high risk in all sites; medium risk in two sites; medium-risk cases with a violation (two sites), and low-risk cases with a violation (one site)]

- At least 1 year of probation remaining

Exclusion criteria included:

- Juveniles
- Non-English speakers
- Transfers and interstate compact
- Special caseloads (e.g., drug court, pretrial, sex offenders)

The original plan was to randomly assign 400 HOPE-eligible individuals in each of the four sites—yielding a total study population of 800 HOPE and 800 PAU probationers. This sample size was sufficient to provide adequate statistical power for hypothesis testing at the site level (as well as greater power overall). Although the enrollment period for the study was extended several times, final enrollment was somewhat less: 1504—743 HOPE and 761 PAU.

Process evaluation/fidelity assessment. This component documented the extent to which each program conformed to the HOPE model; documented the barriers, challenges, facilitators, and lessons learned during implementation to fill gaps in the knowledge base as to what is required to set up a HOPE program; and provided evidence as to the generalizability and sustainability of HOPE programs. This component also assessed implementation fidelity of evidence-based drug treatment programs used by the HOPE programs. Implementation and process measures were collected through ***stakeholder interviews; observation of initial warning hearings and court appearances; and review of court, probation, and HOPE project records***. In addition, detailed data were provided to the evaluation team by the program coordinators employed by the sites and overseen by the training and technical assistance providers led by Angela Hawken at Pepperdine University.

Outcome evaluation. The outcome study assessed whether HOPE participation improved appointment compliance, drug test results, rearrest rates, revocation rates, jail days served, and prison days sentenced. In addition, the evaluation examined whether HOPE participation changed potential mediators including criminal thinking/attitudes, perceptions of control and justice system fairness and legitimacy, dynamic recidivism risk factors, and employment and housing stability. For the outcome study, ***administrative data*** for all four sites provided information on appointment compliance, drug test results, re-arrests, violations, revocations, and jail and prison days for all HOPE and PAU evaluation participants. Data were obtained from local offices (e.g., probation offices, jails, courts), and state agencies (e.g., departments of corrections and probation and parole, state police).[b] ***Interview data*** were collected from evaluation participants at evaluation enrollment (prior to random assignment) and 6- and 12-months post enrollment on measures to facilitate understanding of the nature of individual change associated with HOPE participation. A random sample of study participants who consented and completed baseline interviews were also asked to participate in a twice-weekly series of mini-interviews (Telephone-Audio Computer Assisted Self interviews or T-ACASI technology). Random ***oral swab drug tests*** were collected on a subsample of HOPE and PAU evaluation participants in conjunction with the follow-up interviews (in community only) to provide a common measure of current drug use. When permitted, individuals received modest compensation for completing interviews. All activities were

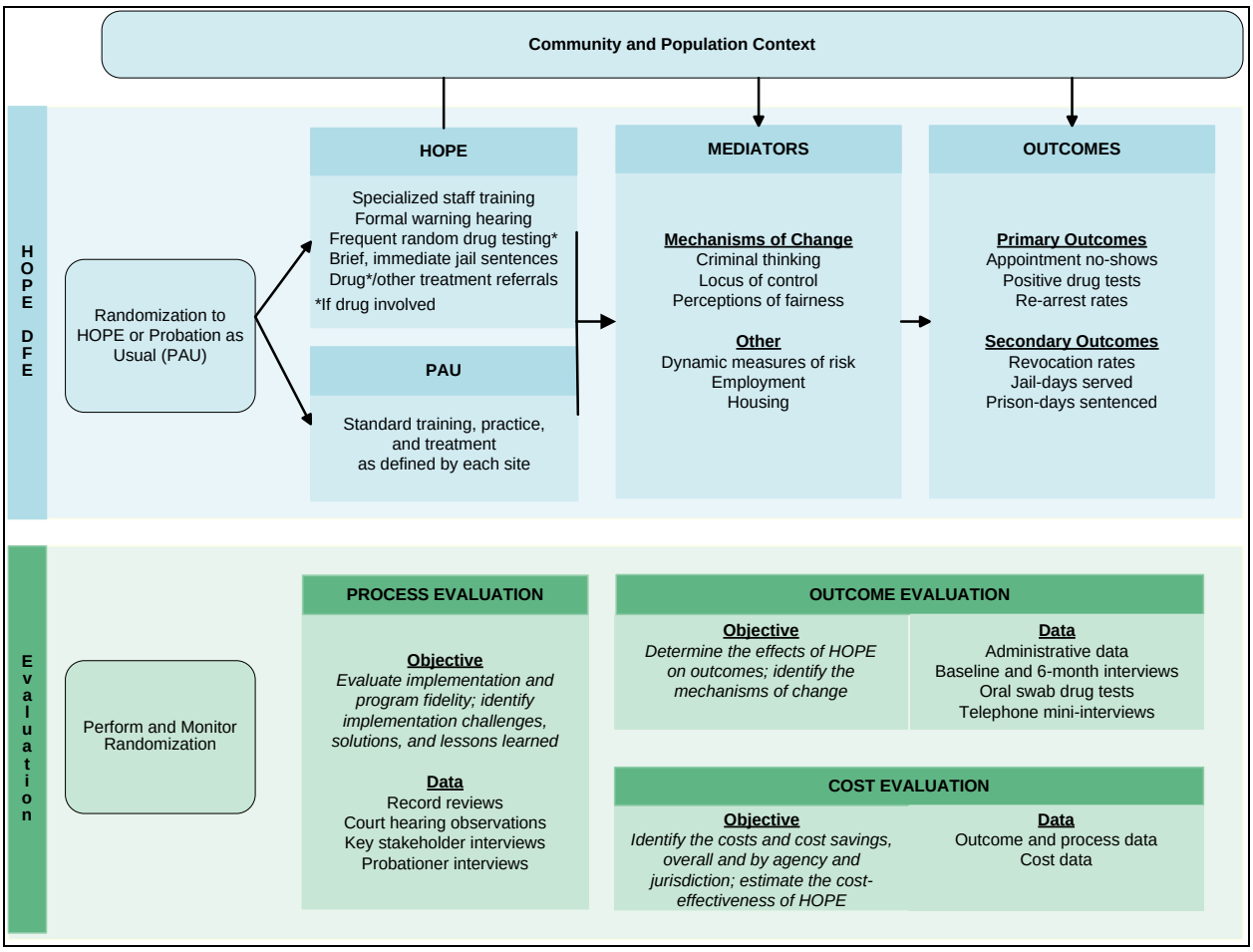
[b]Despite repeated requests, we were unable to obtain arrest data from the FBI National Crime Information Center (NCIC) that would have provided arrest (and in some cases incarceration) records from most states.

under the oversight of RTI's Institutional Review Board; all procedures were submitted to the NIJ Human Subjects Protection Officer for review.

Cost evaluation. The cost evaluation combined data from the process and outcome studies and cost estimates from the local sites and the literature to identify costs and costs savings overall and by jurisdiction. In addition, the cost evaluation assessed the cost-effectiveness of the HOPE program model.

The DFE and evaluation design are shown in ***Exhibit E-1***, which shows random assignment to HOPE or probation as usual (PAU); mediators including mechanisms of change (criminal thinking, locus of control, and perceptions of fairness), dynamic risk factors, employment stability, and housing stability; and the primary and secondary study outcomes (appointment no-shows, positive drug tests, re-arrest rates, revocation rates, jail days served, and prison days sentenced). Evaluation components, objectives, and data sources are also shown.

Exhibit E-1. HOPE DFE model and evaluation design



FINDINGS

Implementation fidelity was assessed against 11 key metrics that were (1) central to the underlying HOPE model (e.g. swift, certain, fair); (2) explicitly set as expectations of DFE sites in the BJA solicitation; (3) within the control of the DFE sites; and (4) measurable with available data. ***Exhibit E-2*** shows the metrics and the degree to which each site met each metric. We present item-level assessments based on whether the site met a standard at least 60% of the time and at least 80% of the time. ***Implementation ranged from good to excellent***.

As can be seen, all sites met at least the 60% standard on all metrics except for holding a violation hearing within 3 days of the violation. All sites struggled to meet this metric and only one site achieved the 60% standard. Achieving this standard was made difficult in some sites by large geographic areas to cover, lack of adequate resources to service warrants, and the challenge in locating absconders. Although the sites missed the 3-day mark, further examination of the data showed that more than three-quarters of all violators did have a hearing within one week (7 days) of the violation.

Implementation fidelity was promoted by several factors:

- A training and technical assistance (TTA) provider was contracted by BJA to provide regular support and guidance to the sites. This provider, Pepperdine University led by Dr. Angela Hawken, was a regular presence at the sites and provided ongoing feedback and correction. Moreover, the provider also sometimes included Judge Alm in these visits.
- Staff from BJA also monitored program implementation.
- Each site had a full-time HOPE project coordinator, paid through the site's BJA grant award, whose role was to assure adherence to the HOPE principles; these individuals also collected the fidelity data for the TTA provider on a real-time basis.

Process Findings. Evaluation team interviews with HOPE stakeholders revealed that, overall, ***there was strong buy-in to the HOPE concept.*** Those implementing the program believed in the model and were optimistic that HOPE would be successful. ***Implementation was facilitated by existing agency cooperation, prior experience with HOPE-like programs, and organizational linkages between probation and the court.*** The local administrative structure of probation was also important to implementation. In three sites, the probation department was either directly under the control of the HOPE judge or there was a sufficiently close administrative linkage such that the judge could substantially direct the operations of probation in the service of the HOPE program—this connection facilitated the establishment of HOPE and its operation. In the fourth site, probation was operated locally from the Sheriff's Department and the HOPE judge had no administrative connection to probation which resulted in challenges. Other challenges in some sites included resource constraints—even with grant funding—and conflicts with existing probation culture (e.g. risk-needs-response versus the surveillance required by HOPE).

POSITIVE VIEWS OF HOPE

HOPE team members had positive views of HOPE believing it was the way probation should be. HOPE team members and HOPE probationers stated that HOPE helped probationers better manage their lives.

Exhibit E-2. Implementation fidelity across the DFE sites

HOPE Fidelity Item	Saline County, AR	Essex County, MA	Clackamas County, OR	Tarrant County, TX
1. Leadership identified by HOPE team members?	83% identified a leader, most commonly the HOPE Judge	83% identified a leader, most commonly Superior Court HOPE Judge	92% identified a leader, no singular leader clearly identified	100% identified a leader, most commonly the HOPE Judge with some secondary endorsement of probation management and the HOPE project coordinator
2. Probationers High Risk	24% of HOPE probationers were moderate to high risk	88% of HOPE probationers were moderate to high risk	80% of HOPE probationers were moderate to high risk	91% of HOPE probationers were moderate to high risk
3. Warning Hearing Compliance with Model Warning Hearing Script	86% of 14 warning hearings complied	95% of 42 warning hearings complied	40% of 5 warning hearings complied	100% of 19 warning hearings complied
4. Initial Drug Testing Frequency	88% of HOPE probationers had at least 8 tests in first 2 months	91% of all HOPE probationers had at least 8 tests in first 2 months	87% of HOPE probationers had at least 8 tests in first 2 months	90% of HOPE probationers had at least 8 tests in first 2 months
5. Stepped Down Drug Testing Frequency	82% of HOPE probationers had at least 1 test per month after first 2 months	83% of all HOPE probationers had at least 1 test per month after first 2 months	73% of HOPE probationers had at least 1 test per month after first 2 months	81% of HOPE probationers had at least 1 test per month after first 2 months
6. Exceptions for Missed Drug Tests	98% of 146 cases with a missed drug test received a consequence	89% of 195 cases with a missed drug test received a consequence	99% of 362 cases with a missed drug test received a consequence	100% of 221 cases with a missed drug test received a consequence
7. Time to Violation Hearing	38% of 639 violations were followed by a violation hearing within 3 days	56% of 736 violations were followed by a violation hearing within 3 days	37% of 1136 violations were followed by a violation hearing within 3 days	66% of 1199 violations were followed by a violation hearing within 3 days
8. Sanction Type	88% of sanctions were jail	78% of sanctions were jail	92% of sanctions were jail	97% of sanctions were jail
9. Sanction Dosage: Jail days <= 19 days (Hawaii HOPE mean)	78% of jail sanctions at or below (Mean = 6 days)	85% of jail sanctions at or below (Mean = 4 days)	65% of jail sanctions at or below (Mean = 6 days)	93% of jail sanctions at or below (Mean = 5 days)
10. Sanction Certainty	97% of violations resulted in a sanction	91% of violations resulted in a sanction	96% of violations resulted in a sanction	>99% of violations resulted in a sanction
11. Sanction Swiftness	76% of sanctions began within 3 days of the violation hearing	60% of sanctions began within 3 days of the violation hearing	63% of sanctions began within 3 days of the violation hearing	83% of sanctions began within 3 days of the violation hearing
SUMMARY	**60% standard: 9 items** **80% standard: 7 items**	**60% standard: 10 items** **80% standard: 8 items**	**60% standard: 9 items** **80% standard: 6 items**	**60% standard: 11 items** **80% standard: 10 items**

Note: Results were similar for the two Massachusetts courts and are combined here.

HOPE providers stated that implementing and operating HOPE was not overly burdensome and that any burden was "worth it." Stakeholders also consistently reported that they believed that HOPE was helping probationers to learn how to better manage their lives more generally through the setting and enforcement of expectations. ***Probationers, who had positive views of HOPE, also stated that they believed that HOPE had helped them better manage their lives***. HOPE team members and probationers noted that the frequent drug testing regimen could create conflicts for probationers who were employed and who lived (or worked) distant from the testing site. In some cases, probationers lost jobs because of the testing requirements.

HOPE team members also expressed some concerns that HOPE did not work as well for lower functioning or mentally ill probationers and for the more seriously antisocial probationers who were not threatened by a few days in jail . This latter view was confirmed by some of the interviewed probationers who reported that they were just going along with the program until their probation ended at which time they would go back to their antisocial lifestyles.

HOPE probationers understood what was expected of them. Both study groups had a strong sense that their probation officer would find out about noncompliance and would arrest them or have them arrested for noncompliance. Both groups also had a strong sense that the judge would do something in response to noncompliance, although HOPE probationers at their 12-month interview were more certain than PAU probationers that the judge would respond suggesting that the HOPE probationers—if they didn't understand initially—learned that sanctions would happen.

HOPE probationers who participated in a qualitative interview were mixed as to whether they thought about the potential consequences before committing violations. Most did not report giving the possibility of punishment much thought (even though they knew in an intellectual sense what could happen), with some suggesting that they did not actually care much about being punished. For some, though, the deterrence message set in over time, leading them to be more thoughtful about their behavior. Responses to ACASI interviews with probationers underscore this point: ***HOPE probationers were more sensitive to the possible consequences of noncompliance (as measured by the deterrence score) and reported a lower tolerance for law violations than their PAU counterparts.*** The ACASI interviews also offer some evidence about change in other attitudes among HOPE probationers. Specifically, at follow-up, ***HOPE probationers reported greater self-efficacy and a lower level of identification with crime-involved people than PAU probationers***.

ACASI interviews—and our extensive transition analyses—show that ***HOPE and PAU probationers experienced probation differently.*** Although HOPE and PAU probationers were equally likely to be required to attend substance abuse treatment as a condition of supervision, HOPE probationers were more likely to attend treatment. More HOPE probationers than PAU probationers were subjected to drug testing as a supervision requirement and very few PAU probationers were subjected to random testing. HOPE probationers who participated in a qualitative interview felt that ***the most difficult part of HOPE was balancing the need to report for frequent drug tests with their work schedules***, leading some to lose jobs due to their participation in HOPE. ***They also felt that HOPE's emphasis on accountability was helpful, as was the structure it provided,*** which was often lacking before HOPE.

A final consideration is that some HOPE probationers came to rely on the drug testing hotline to maintain their sobriety and were afraid that once they were no longer tested that they would no longer be able to maintain sobriety. A similar finding was also reported by Hawken and Kleiman (2009).

Exhibit E-3. HOPE Evaluation Enrollment

	HOPE	PAU	Total
AR	179	163	342
MA	189	203	392
OR	190	204	394
TX	185	191	376
Total	743	761	1504

Outcome Findings. A total of 1,580 individuals were randomly assigned to HOPE (794) or to PAU (786) between August 2012 and September 2014[3] (***Exhibit E-3***). Of these, 76 individuals were determined after random assignment to be study ineligible (68 individuals were program ineligible; and 8 were randomized twice and were retained in their original study and program assignment). Most were male (81%), white (69%), and high risk (55%). On average, they were 31 years at study enrollment, with 7 prior arrests and 3.5 prior convictions (***Exhibit E-4***). Most were on probation for either a drug (31%) or property offense (30%).

Exhibit E-4. Characteristics of subjects enrolled in the DFE, overall and by site

	Overall	AR	MA	OR	TX
Age at intake* **	31.1 (10.4)	32.3 (10.2)	33.7 (11.1)	30.8 (9.9)	**27.5† (9.1)**
Male = 1* **	0.81 (0.40)	0.73 (0.44)	0.88 (0.32)	0.83 (0.38)	0.77 (0.42)
Race = White* **	0.69 (0.46)	0.85 (0.36)	0.68 (0.47)	0.88 (0.33)	0.35 (0.48)
High risk* **	0.55 (0.50)	0.03 (0.17)	0.72 (0.45)	**0.88† (0.33)**	0.52 (0.50)
Age at first arrest* **	22.1 (7.78)	27.2 (9.63)	20.0 (6.53)	22.9 (6.17)	19.0 (5.95)
# Prior arrests* **	7.30 (8.13)	4.4 (3.38)	13.0 (11.9)	6.0 (6.05)	5.4 (4.70)
Prior person charge* **	0.56 (0.50)	0.52 (0.50)	0.86 (0.35)	**0.48† (0.50)**	**0.38† (0.49)**
Prior property charge* **	0.74 (0.55)	0.69 (0.46)	0.81 (0.39)	0.69 (0.46)	0.77 (0.42)
Prior drug charge* **	0.66 (0.48)	0.59 (0.49)	0.57 (0.50)	**0.73† (0.44)**	0.73 (0.44)
Prior public order/other	0.77 (0.42)	0.74 (0.44)	0.93 (0.26)	**0.76† (0.43)**	0.65 (0.48)
# Prior convictions* **	3.54 (4.42)	1.7 (1.13)	5.8 (6.45)	4.3 (3.91)	2.1 (2.51)
Study Offense					
Person* **	**0.24† (0.42)**	0.23 (0.42)	0.50 (0.50)	0.20 (0.40)	0.01 (0.10)
Property* **	0.30 (0.46)	0.37 (0.48)	0.22 (0.42)	0.16 (0.36)	0.48 (0.50)
Drug* **	0.31 (0.46)	0.24 (0.43)	0.11 (0.31)	0.44 (0.50)	0.44 (0.50)
Public order/other	**0.15† (0.36)**	0.16 90.36)	0.17 (0.38)	**0.20† (0.40)**	0.08 (0.26)
N	1504	342	392	394	376

***Subject characteristics differ across sites ($p < 0.001$).

†HOPE and PAU differ at $p < 0.05$ (***see Exhibits 3-7*** through ***3-11***.

[3] Enrollment was August 2012 through December 2013 (Oregon), August 2012 through September 2014 (Arkansas and Texas), and October 2012 through July 2014 (Massachusetts). The delay in program start-up in Massachusetts was due to a statewide hiring freeze; enrollment continued until 400 HOPE-eligible cases were identified and randomly assigned (Massachusetts and Oregon) or until enrollment was ended somewhat short of the 400 site-level goal in September 2014 (Arkansas and Texas)

Subject characteristics varied across the sites. For example, study participants were younger at first arrest in Texas than Arkansas (19 versus 27 years) and the average number of prior convictions ranged from 1.7 in Arkansas to 5.8 in Massachusetts.

Interviews were conducted at baseline and at 6 and 12 months following baseline. ***Exhibit E-5*** summarizes interview completions by wave and site. Response bias analyses suggested no differences between those who were interviewed and those who were not interviewed at any wave(Section 2.3).

Exhibit E-5. HOPE interview completion summary counts

Interview outcomes	AR	OR	TX	MA	Total
Baseline completes	170	245	217	346	978
Baseline response rates (%)	50	62	58	88	65
6-month completes	145	128	91	172	536
6-month response rates (%)	42	32	24	44	36
12-month completes	137	116	88	118	459
12-month response rates (%)	54	29	26	36	35
Total N	342	392	394	376	1504

Interview findings showed the following:

- Probationers on HOPE and PAU reported similar rates of employment and similar wages across the three waves of interviews.
- HOPE probationers were more likely to have a job with formal pay at the 12-month interview.
- Both groups reported emotional problems. Overall HOPE probationers reported a lower average mental health symptom level at 12 months. There were no differences at any interview wave in self-assessment of emotional problems interfering with work or other activities. About 40% of both groups reported at all waves that they had accomplished less than they would have liked because of emotional problems.
- Both groups agreed that they needed education and were neutral with respect to needing job training.
- Neither group believed that they needed substance abuse or mental health treatment.
- There were no differences between the groups on the receipt of education and employment services.
- HOPE probationers were more likely than PAU probationers to report having received residential substance abuse treatment at the 6- and 12-month interviews.
- At 12-month interview, HOPE probationers were much less likely than PAU probationers to report that most or all close friends are frequently drunk or high (10% versus 20%) or that most or all close friends have been incarcerated (20% versus 31%).

HOPE was to hold individuals accountable to their supervision conditions, including compliance with intensive random drug testing—suggesting that HOPE probationers would have more violations which is what was observed. HOPE probationers were more likely to have a violation (89% versus 82%) and had more violations than PAU probationers (3,770 versus 3,134)—mostly drug-related violations (2,107

versus 915) attributable to testing (26,991 for HOPE versus 4,942 tests for PAU probationers). ***HOPE probationers were less likely to miss a probation officer visit (30% versus 44%), to fail to pay their fees and fines (11% versus 18%), and to be violated for a new charge (22% versus 28%). HOPE probationers were more likely to have a violation for failing to appear for court (18% versus 6%), although this may be because they had more hearings***. Most sanctions for HOPE probationers were jail (2,920 of 3,550 sanctions) and HOPE probationers were more likely to go to jail (82% versus 56%), to go more often (3.8 stays versus 1.4), and to serve more days total (47 versus 33.3 days) with a median stay of 4 days.

HOPE included treatment referral after repeated failed tests; HOPE participants were more likely to go to residential treatment (33% versus 11%). HOPE probationers were also referred to treatment more quickly (overall and in three sites). ***Drug tests conducted in conjunction with follow-up interviews showed fewer positives for HOPE than PAU probationers***.

Recidivism outcomes were similar for the HOPE and PAU groups: 40% of HOPE versus 44% of PAU had a new arrest; 25% of HOPE versus 22% of PAU had a revocation; 49% of HOPE versus 50% of PAU had an arrest or revocation; and 28% of HOPE versus 26% of PAU had a new conviction. There was some variation in rates across sites, but the general conclusions of no differences hold with two exceptions: (1) HOPE probationers were more likely to be revoked in two sites (PAU revocation rates in those sites were low—about 10%.); and (2) HOPE probationers were more likely to have a new conviction in one site. Lognormal survival models of time to recidivism events (***Exhibit E-6***) confirm the bivariate findings, but revealed one additional finding. Time to revocation was less for HOPE probationers in Arkansas and Oregon, while HOPE probationers had longer times to revocation in Texas.

Exhibit E-6. Lognormal survival model results for time to recidivism events

Outcome	Parameter Estimates (HOPE 1) and (Standard Errors)				
	Overall[1]	Arkansas	Mass.	Oregon	Texas
Rearrest	0.14 (0.13)	0.12 (0.27)	0.11 (0.19)	0.12 (0.30)	0.21 (0.27)
Revocation	-0.17 (0.11)	**-1.02* (0.23)**	0.17 (0.23)	**-0.94* (0.46)**	**0.41* (0.14)**
Rev/Rearrest	0.03 (0.11)	-0.24 (0.24)	0.14 (0.18)	-0.08 (0.26)	0.21 (0.19)
Reconviction	-0.07 (0.12)	**-0.88* (0.28)**	0.06 (0.21)	0.15 (0.18)	0.27 (0.29)

[1]Overall models included controls for site

*p < 0.05

Cost findings. Cost analyses estimated costs of intake, warning hearings, staffing meetings, office visits, drug tests, violation hearnings, arrests, state and county corrections, and residential treatment. ***Six-month median costs were significantly higher for HOPE than PAU overall and in four sites; mean costs were higher overall and in three sites (Exhibit E-7***).

Exhibit E-7. Average total costs per probationer by site, 6-month sample

Site	PAU			HOPE			t test	Median test
	N	Mean	Median	N	Mean	Median		
AR	**159**	**$892**	**$190**	**179**	**$1,893**	**$1,038**	**5.4*** **	**56.5*** **
MA	199	$2,128	**$329**	188	$2,223	**$651**	0.3	**30.7*** **
OR	203	**$2,836**	$1,162	190	**$3,562**	$1,983	**2.0***	3.5
TX	191	**$1,813**	**$639**	185	**$3,913**	**$2,723**	**7.3*** **	**47.8*** **
All	*752*	***$1,978***	***$364***	*742*	***$2,908***	***$1,698***	***5.8*** **	***84.8*** **

*p < 0.05; **p < 0.01; *** p < 0.001

Exhibit E-8 shows that 12-month median and mean costs were significantly higher overall and in three sites.

Exhibit E-8. Average total costs per probationer by site, 12-month sample.

Site	PAU			HOPE			t test	Median test
	N	Mean	Median	N	Mean	Median		
AR	112	**$2,420**	**$418**	134	**$4,028**	**$1,939**	**3.0*** **	**4.2***
MA	167	$5,637	**$733**	153	$6,984	**$2,104**	1.3	**6.6***
OR	203	$6,333	$3,826	189	$8,566	$5,044	**2.9*** **	3.0
TX	167	**$4,700**	**$2,150**	166	**$8,386**	**$8,718**	**6.1*** **	**26.0*** **
All	*649*	***$5,059***	***$1,863***	*642*	***$7,195***	***$4,015***	***5.4*** **	***36.5*** **

*p < 0.05; **p < 0.01; *** p < 0.001

Twenty-four-month median and mean costs were significantly higher overall and in one site (Exhibit E-9). For each estimation period, cost differences were driven by treatment and incarceration costs.

Exhibit E-9. Average total costs for 24 months of supervision per probationer by site.

Site	PAU			HOPE			t test	Median Test
	N	Mean	Median	N	Mean	Median		
AR	54	$6,563	$2,139	68	$7,901	$3,184	0.7	0.1
MA	61	$13,425	$2,721	56	$17,672	$6,727	1.0	3.8
OR	103	$14,588	$9,600	96	$17,564	$14,015	1.4	3.1
TX	93	**$9,392**	**$5,261**	94	**$15,038**	**$13,799**	**4.1*** **	**4.5***
All	*311*	***$11,413***	***$5,797***	*314*	***$14,735***	***$10,355***	***2.7*** **	***7.6*** **

*p < 0.01**p < 0.01; *** p < 0.001

CONCLUSIONS

Four sites that differed in organizational structures and populations successfully implemented HOPE programs—holding probationers accountable to their conditions of supervision and reducing drug use. ***HOPE was effective in increasing compliance with some supervision conditions (e.g., probation officer visits and payment of fees and fines). HOPE probation also appeared to have positive effects on drug use based on oral swab drug tests conducted in conjunction with interviews 6 and 12 months after program enrollment.*** HOPE probationers were randomly tested at a very high rate (26,991 tests compared to 4,942 for the PAU probationers). This extensive testing led to more violations—as would have been expected—although positive tests reduced substantially over time (again suggesting positive impacts on drug use).

HOPE probationers were more likely to go to jail (82% versus 56%), to have more jail stays (3.8 versus 1.4), and to spend more days in total in jail (47 versus 33 days). HOPE probationers were also more likely to be sent to residential treatment (overall and in three sites).

Overall, ***HOPE did not reduce recidivism***, as measured by arrest, revocation, and new conviction. In two sites, revocations were higher for HOPE than PAU and in one site reconvictions were higher. The sole signficant positive recidivism finding was a longer time to revocation in one site (although revocation rates were similar).

More jail days, more residential treatment, and similar (or higher) recidivism resulted in higher (although not always significantly higher) costs for HOPE compared with PAU probation.

IMPLICATIONS FOR POLICY AND PRACTICE

HOPE probation has been widely promoted and adapted as a means for substantially improving probation outcomes while generating cost savings. The findings of this rigorous four-site randomized controlled trial suggest otherwise. So what is to be made of this?

The results do <u>not</u> say do not implement HOPE or similar programs based on "swift, certain, and fair" principles. The results <u>do</u> suggest that sites considering implementing such programs should give great consideration to the implications of HOPE programs within the context of their current probation policy and practice—PAU context is important.

For example, within the DFE:

- PAU revocation rates were low (9% and 13%) in two sites—suggesting limited ability to reduce revocations and that sites with low revocation rates should consider whether to implement procedures to mitigate any potential increases in revocations that would accompany the increased surveillance associated with HOPE.
- In at least two sites, revocation could yield only short prison stays (up to 90 days)—suggesting limited opportunities for "prison bed savings" even if revocations were lower with HOPE and providing a smaller incentive for individuals to comply.
- PAU was based at least somewhat on Risk-Needs-Response principles in at least two sites—suggesting an additional consideration with respect to the integration of HOPE with PAU, particularly if HOPE supervision requirements would take resources away from RNR activities.
- In one site, probation could use short jail stays on their authority (and did for PAU cases)—suggesting that a HOPE judge was not necessary to enforce conditions.

Thus, the similar recidivism outcomes may hinge on the "compared to what" aspect of any evalution—in that findings suggest that HOPE worked as well as but not better than PAU. However, given the consistency of findings across four sites that differed in the administration of PAU, there is little to support a conclusion that HOPE or HOPE-like programs will produce substantial improvements over PAU when implemented widely. This may be particularly true given the extensive resources that were devoted to facilitating implementation in the DFE—resources likely much greater than sites would generally have to implement a new program.

Additional research is needed to determine whether there may be specific types of probationers who would be more responsive to this type of program. For example, perhaps there are subpopulations for whom the threat of even short jail stays may provide more motivation. Perhaps HOPE is most suitable for more malleable "mid-range" offenders—those who need some intervention but who are not so deeply entrenched in a criminal lifestyle that the threat of sanctioning alone is insufficient to elicit behavioral change. At least anecdotally it also may be that HOPE provides a useful crutch for those who have decided to leave drugs and a criminal lifestyle behind. We heard from multiple individuals that there were HOPE probationers who wanted to remain on the random drug testing regime because it was helping them stay sober. In contrast, we heard from others more entrenched in an antisocial lifestyle who either didn't care about the HOPE sanctions or stated that they were complying to get through the program with every intent to return to a drugs and crime lifestyle after HOPE probation ended. More research is needed to develop our understanding of who (if anyone) is likely to maximally benefit from a swift and certain sanctioning regimen such as HOPE.

In a recent discussion with Judge Alm (personal communication following the initial release of findings), he stressed that the Hawaii HOPE model was predicated on layering the swift, certain, and fair sanctions on top of an RNR, motivational interviewing (MI) approach to probation supervision. (Also see Alm, 2016.) This is certainly not the way that the program was described in the BJA solicitation that sought DFE implementation sites (https://www.bja.gov/Funding/11HOPEsol.pdf). Nor is it the way the program and SCF approaches have been described more generally over the past several years. It is possible that RNR/MI/SCF supervision may be more successful than other approaches to probation. This model was not explicitly tested by this evaluation, although the fact that two of the sites followed RNR principles as PAU suggests that (1) there may have been a test although the RNR component was not measured and at least one of the sites indicated that they struggled to follow RNR principles while complying with the other demands on probation officer time that accompanied HOPE implementation; and (2) HOPE in those sites was being compared to RNR—which is promoted as a best practice for probation. Although our fidelity assessment was thorough in measuring adherence to HOPE principles, these principles were those specified by BJA in their program solicitation and did not include any mention of RNR or MI as components of HOPE. Hence, these were not measured so it is unknown the extent to which PAU and HOPE in these sites used RNR or MI in their supervision.

It is also important to consider that part of the strong appeal of HOPE is the relative simplicity of implementing and the salience of assumed effectiveness of its strict supervision and sanctions-based approach. These factors, coupled with claims that HOPE programs would save money, are likely reasons that many jurisdictions were quick to adopt the approach. HOPE appeared to be the elusive "silver bullet" that would easily dispatch bad behaviors. That "hope" is dispatched by the findings reported here.

Finally, it should be noted that there are others who questioned the potential of HOPE from the beginning (e.g., Clear and Frost, 2014; Cullen, Manchak, and Duriez, 2014) and have subsequently pointed to the findings herein as a serious caution flag for agencies and jurisdictions that continue to embrace HOPE models of supervision (Cullen, Pratt, and Turanovic, 2016; Cullen, Pratt, Turanovic, and Butler, 2018). Cullen et al. (2016:1221) noted that "In the end, we should not allow the language of "hope" and "swift-certain-fair" to cloud what Project HOPE really proposes: zero-tolerance supervision…. Even if Project HOPE worked—which does not appear to be the case—we would not wish to look ahead to a correctional future in which probation and parole were turned into another form of policing aimed at constantly tightening the screws on offenders under supervision." More recently, Cullen and colleagues have provided additional commentary on the status of the implementation of HOPE and SCF programs, reiterating that "…HOPE lacks consistent empirical support and thus is likely to be yet another in a long line of deterrence-oriented interventions that, in the end, offers false hope [Duriez et al., 2014])." (Cullen et al., 2018: 29).

1. Introduction

In September 2011, the National Institute of Justice (NIJ) awarded RTI International and the Pennsylvania State University Justice Center for Research (RTI/PSU) a grant (NIJ 2011-RY-BX-0003) to evaluate the Honest Opportunity Probation with Enforcement Demonstration Field Experiment (the HOPE DFE Evaluation). The HOPE DFE Evaluation was implemented to determine whether a supervision program originally implemented in Hawaii could be replicated in multiple sites in the continental United States and to provide a rigorous, independent test of HOPE's capacity to yield substantial reductions in appointment no-shows, positive drug tests, re-arrests, revocations, and jail and prison days. The evaluation included a process evaluation, an outcome evaluation, and an economic evaluation.

A critical element of the evaluation design was the implementation of a rigorous randomized controlled trial (RCT) in each of four sites that implemented HOPE programs with grant funds and support from the Bureau of Justice Assistance (BJA)[4]:

- Saline County/Benton, Arkansas (AR)
- Essex County/Salem, Massachusetts (MA)
- Clackamas County/ Oregon City, Oregon (OR)
- Tarrant County/Fort Worth, Texas (TX)

This section provides an overview of the HOPE program model and history, introduces the evaluation design, and provides information on the registration of the RCT.

EVALUATION GOALS

1. Design & implement a randomization process in each of four DFE sites.

2. Conduct a comprehensive process evaluation to assess implementation fidelity and identify lessons learned to enhance future replications.

3. Conduct a rigorous experimental outcome evaluation to determine the effect of the HOPE model on individual probation outcomes.

4. Conduct a cost evaluation to assess the cost effectiveness of HOPE.

1.1. The HOPE Supervision Model

The HOPE supervision model was developed in 2004 as Hawaii's Opportunity Probation with Enforcement (Hawaii HOPE) program, building on similar efforts that use certain, but non-severe, graduated sanctions to deter probationers from violating supervision conditions. Hawaii HOPE was developed under the direction of Judge Steven Alm and emphasized close monitoring, frequent random drug testing, and "swift, certain, and fair" (SCF) sanctioning, reserving scarce treatment resources for those most in need. The HOPE model contrasts with more traditional approaches to probation in which multiple violations are tolerated until "a last straw" results in revocation—often to a lengthy prison term. As Judge Alm (2015) noted, "Where probation-as-usual is often delayed, uncertain, inconsistent, and, when court action is finally taken, often unnecessarily harsh, HOPE is swift, certain, consistent, and proportionate."

[4] The solicitation for proposals to implement the HOPE programs for the DFE is here: https://www.bja.gov/Funding/11HOPEsol.pdf.

Thus, the principles of HOPE—swiftness, certainty, consistency, and proportionality/fairness—were posited to address potential conclusions a probationer could draw from the operations of probation as usual (PAU):

- Tolerance for multiple violations suggests that conditions are not important and that sanctions are unlikely;
- "Draconian" response to a single violation (i.e., lengthy prison term) suggests that punishment is "unfair;"
- Tolerance for violations until there are "one too many" suggests that sanctions are "random" and disconnected from the probationer's behavior; and
- Temporal disconnects between violations and sanctions suggest that probationers don't link the two.

Indeed, the intrusiveness of conditions, probation officers' knowledge of misbehavior, and probation officer response to misbehavior have been found to not affect criminal activity or violations of probation conditions in a traditional probation setting (MacKenzie, Browning, Skroban, and Smith, 1999). Further, traditional probation is counter to what is known scientifically about shaping human behavior (e.g., Harrell and Roman, 2001; Kennedy, 1997).

Exhibit 1.1 illustrates the HOPE model. Individuals assigned to HOPE supervision are issued a warning hearing by the HOPE judge at the time they are placed on HOPE supervision. During the hearing, the judge informs the probationer that they will be held accountable for complying with all supervision conditions, will be subject to frequent random drug testing, and will be sanctioned for any violation, including any positive or missed drug tests. The warning hearing follows a script to assure that all key points are addressed.

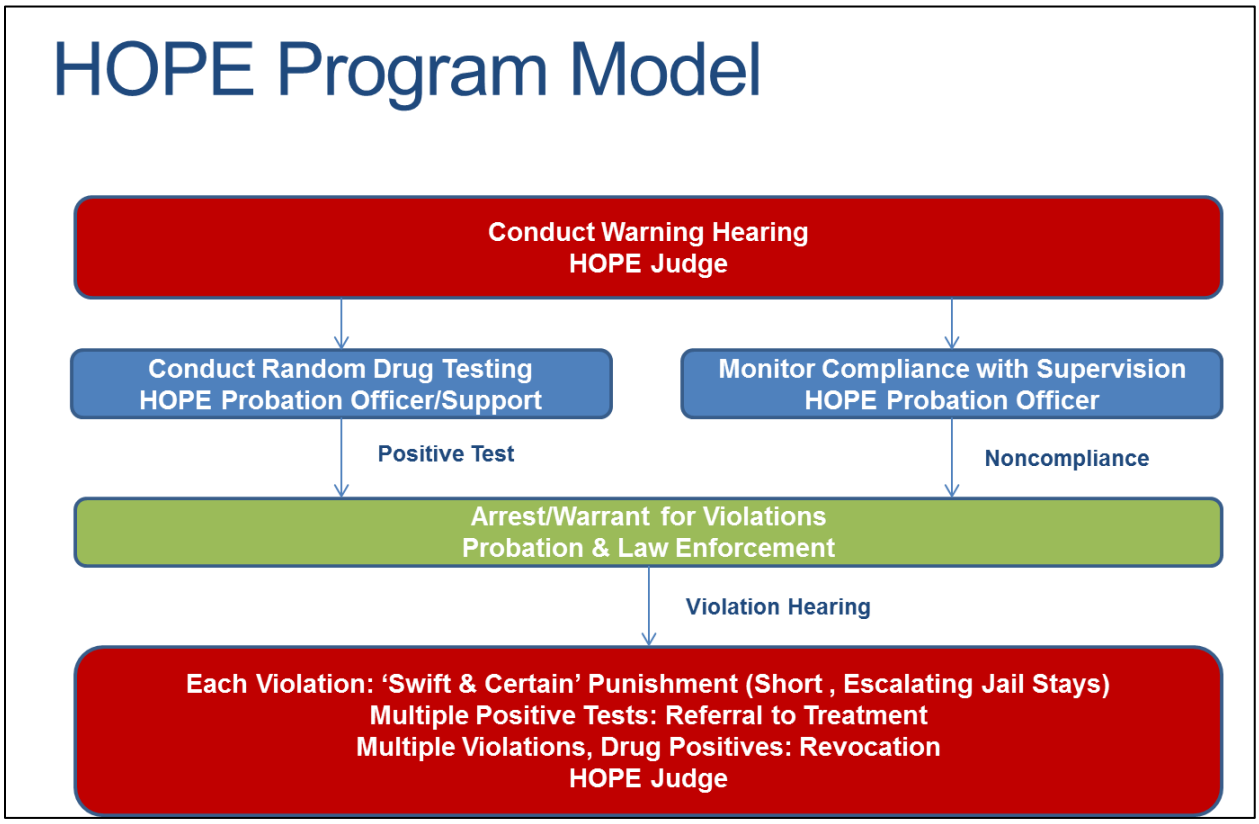
All violations are (generally) followed by an arrest or warrant and the individual is brought "immediately" before the judge for a violation hearing. Each violation is met with a punishment—generally a few days in jail to be served immediately although in some cases judges choose to delay jail until the weekend (to accommodate employment) or impose other sanctions such as community supervision. With HOPE, the *certainty* of punishment is combined with *short* jail stays—hence, the most recent framing of this approach to supervision as swift, certain, and fair (SCF; see scfcenter.org/ or swiftcertainfair.com/). Repeated violations result in escalating sanctions and repeated positive drug tests result in referral to treatment. Thus, under the HOPE model, scarce treatment resources, shown to be effective in reducing criminal behavior (e.g., Lattimore, Krebs, Koetse, Lindquist, and Cowell, 2005; MacKenzie, 2006), are reserved for those most in need.

More recently, Judge Alm has stressed that the Hawaii HOPE model was predicated on layering the swift, certain, and fair sanctions on top of an Risk-Needs-Response (RNR), motivational interviewing (MI) approach to probation supervision (e.g., Alm, 2016.) This is certainly not the way that the program was described in the BJA solicitation that sought DFE implementation sites (https://www.bja.gov/Funding/11HOPEsol.pdf). Nor is it the way the program and SCF approaches have been described more generally over the past several years.

Finally, it should be noted that others questioned the potential of HOPE from the beginning (e.g., Clear and Frost, 2014; Cullen, Manchak, and Duriez, 2014). More recently, Cullen and colleagues have provided additional commentary on the status of the implementation of HOPE and SCF programs,

reiterating that "…HOPE lacks consistent empirical support and thus is likely to be yet another in a long line of deterrence-oriented interventions that, in the end, offers false hope [Duriez et al., 2014])." (Cullen et al., 2018: 29).

Exhibit 1-1. HOPE supervision program model



Theoretical Underpinnings

The logic model underlying HOPE supervision presumes that if each violation is responded to with a violation hearing and sanction the probationer will be *deterred* from violating conditions of supervision and engaging in new criminal activities. The concept that criminals engage in a decision-making calculus that explicitly weights rewards and punishments dates to Beccaria (1963 [1764]) and Bentham (1789). Becker (1968) proposed an economic model of criminal choice based on expected utility theory (also see Block & Heineke, 1975; Ehrlich, 1973; Witte, 1980). Theories of crime based on classic economic models of choice were subsequently criticized for failing to capture cognitive activities involved in criminal behavior (Cornish & Clarke, 1986). Indeed, the adequacy of the expected utility model more broadly to explain choice under uncertainty was criticized and its assumptions empirically refuted by Kahneman and Tversky (e.g., 1979, 1982), who proposed prospect theory as an alternative.

A prospect theory model of criminal choice was proposed by Lattimore and Witte (1986) and tested along with alternative formulations based on expected utility and subjective expected utility theory (Lattimore, Baker, & Witte, 1992). For their study, certainty equivalent data were elicited for risky gains and losses from male and female undergraduate college students and incarcerated 18-to-22–year-old

property offenders. The choice scenarios were potential gains from a breaking and entering crime (e.g., 50% chance of $50 and a 50% chance of $0 dollars), and potential losses of freedom associated with a plea bargain versus a trial (e.g., 50% chance of probation and 50% chance of 12 months in jail). Findings from the study provided support for decision making that allows for the subjective weighting of both probabilities and the value or utility of outcomes such as those consistent with prospect theory or subjective expected utility theory.

Recently, there has been a resurgence of interest in rational choice theories of criminal behavior (Apel, 2013; Carmichael and Piquero, 2004; Nagin, 2007, 2013; Paternoster, 2010). Nagin has argued for criminal decision-making theory rooted in the judgment and decision-making research and literature in psychology, including prospect theory. Nagin and Paternoster (1993) provided an empirical test of the role of cost and benefit considerations in criminal choice (theft, drinking and driving, and sexual assault) through analyses of data from a survey of college undergraduates queried on their likelihood of committing each of these offenses under a variety of scenarios. Their findings showed that most respondents reported that they would never commit theft or sexual assault under any scenario (63% and 85%, respectively), and 33% reported that they would never drink and drive. For the others, however, they summarized, "we found that perceptions of the certainty of formal and informal sanctions and self-imposed shame effectively controlled respondents' intentions to offend" (Nagin and Paternoster, 1993: 489).

Other empirical studies (e.g., Wright, Caspi, Moffitt, and Paternoster, 2004; Matsueda, Kreager, and Huizinga, 2006) also provide support for deterrent effects. Finally, Durlauf and Nagin (2011) argue that the accumulating evidence supports a move toward "certainty-based as opposed to severity-based sanction policies" (p. 14; also, see Apel and Nagin, 2011) and conclude that if certainty of sanctions is more important than the severity of sanctions, it may be possible to reduce both crime and the costs of punishment.

It is within this theoretical and empirical framework that HOPE posits that many offenders will choose not to use drugs if they are certain to be caught and sanctioned. The underlying framework of HOPE is to allow probationers to learn the relationship between their behavior and official response to that behavior: violations will be met with sanctions, even if the severity of the sanctions is low. This sanctioning approach incorporates deterrence, as well as conditioning and learning theories, to teach probationers that violations have consequences and should result in changes in attitudes, perceptions of individual control over consequences, fairness, and legitimacy.

If HOPE supervision results in reductions in revocations and new criminal behavior, the program is also anticipated to reduce criminal justice system costs—producing savings in prison bed days that will more than offset the costs of warning hearings, additional violation hearings, short stays in jail, warrant service, arrests, and drug testing. At a minimum, HOPE is assumed not to impose additional public safety costs in the form of increased criminal activity over supervision as usual.

Evidence for the HOPE Model

The original studies of Hawaii HOPE included a quasi-experimental pilot evaluation and a subsequent study in which probationers deemed "high risk" by their probation officers were randomly assigned to either HOPE or PAU. This latter study (Hawken and Kleiman, 2009) found significant improvements for HOPE participants at 1-year follow-up:

- 14 percentage point reductions in missed probation appointments,
- 33 percentage point reductions in positive urine tests,
- 26 percentage point reductions in new arrests, and
- 8 percentage point reductions in probation revocations.

More recently, Hawken and colleagues (2016) reported the results of a 76-month follow-up of the original HOPE experimental subjects that show less dramatic long-term effects with no significant differences in the percentage experiencing a new charge (42% of HOPE probationers and 47% of those on PAU). The average number of new charges was less for HOPE than for control subjects (0.91 new charges compared with 1.12), although most of the difference was a result of a reduction in the number of new drug charges (0.12 compared with 0.27). Revocations were reduced for HOPE probationers (13% vs. 27%), and HOPE probationers were more likely to experience early termination of probation—both of which resulted in criminal justice system savings.

HOPE/SCF builds on earlier efforts such as Project Sentry that employed random drug testing of released jail inmates and imposed immediate sanctions for noncompliance (Buntin, 2009; Hawken and Kleiman, 2007; Project Sentry, 2004). Hawken et al. (2016) summarized results from recent evaluations of HOPE-like programs. Quasi-experimental evaluations (matched comparison, pre–post design) suggested strong positive results from the Supervision with Intensive Enforcement (SWIFT) program in Tarrant County, Texas (Snell, 2007), whereas results were mixed for the Alaska Probation Accountability and Certain Enforcement (PACE) program that showed reduced drug positives and missed appointments but increased violations (Carns and Martin, 2011).

Washington State adopted a Swift and Certain (SAC) policy statewide with a goal of reducing prison time for those violating conditions of community supervision. Hamilton, van Wormer, Kigerl, Campbell, and Posey (2015; also, see Hamilton, Campbell, van Wormer, Kigerl, and Posey, 2016). Quasi-experimental findings from the SAC program evaluation support positive effects including reductions in prison confinement and duration resulting from reduced revocations and reduced odds of convictions for those under the new SAC supervision regime in comparison with similar individuals who were supervised prior to the implementation of the SAC policy. Nevertheless, these findings must be interpreted with caution because marijuana was legalized in Washington just as SAC was implemented. Specifically, the comparison group was identified between September 2010 and February 2011 and the SAC group was identified between September 2012 and February 2013. Both groups were followed for 1 year. Adult possession of marijuana was legalized December 2012. Thus, the two groups were exposed to different legal environments, which may account for their findings.

Finally, recent efforts to implement a deterrence-based program to manage high-risk, substance-using probationers in Delaware found that "judicial practices, client eligibility, logistics, and cooperation with secure facilities all posed noteworthy issues for program implementation" (O'Connell, Visher, Martin, Parker, and Brent, 2011: 261). An experimental evaluation of this program entailed random assignment to the program or to probation as usual. Results from this evaluation revealed that the program did not reduce substance use or crime for program participants (O'Connell, Brent, and Visher, 2016).

On the promise of the findings from the original Hawaii HOPE program, multiple states have launched HOPE-like programs (e.g., Petranik, 2011). For example, Hawken and colleagues (2016: 24)

noted that as of January 2015, HOPE or SCF programs had been implemented in 28 states, one Indian nation, and one Canadian province. The HOPE approach has benefited from a charismatic advocate in Judge Steven Alm, who introduced the program in Hawaii; strong advocacy by the researchers who originally studied the Hawaii program; and support by the National Institute of Justice and the Bureau of Justice Assistance (Duriez, Cullen, and Manchak, 2014). Nonetheless, the evidence base for HOPE remains limited (see, for example, Cullen, Pratt, Turanovic, 2016). Replication of the program and, most importantly, additional controlled trial tests of HOPE effectiveness are needed before practitioners can be confident that investing in the HOPE model offers the most cost-effective return on scarce funds. Furthermore, it is important to assess whether a program founded in Hawaii can be generalized to the contiguous United States. Differences between the legal and community contexts and the probation population in Hawaii may account for some of the remarkable early findings from the Hawaii HOPE evaluation.

1.2. Evaluation Design Summary

The HOPE DFE evaluation included a process evaluation to assess program fidelity and implementation challenges and successes; an outcome evaluation to determine the effectiveness of HOPE supervision compared with PAU practices in each of the study sites; and a cost/economic evaluation to assess the costs of HOPE supervision compared to PAU. The outcome evaluation was based on randomized controlled trials in the four sites to assure that the HOPE and PAU groups had similar characteristics so that any differences in observed outcomes could be ascribed to the HOPE intervention.

The evaluation was designed to provide answers to the following research questions:

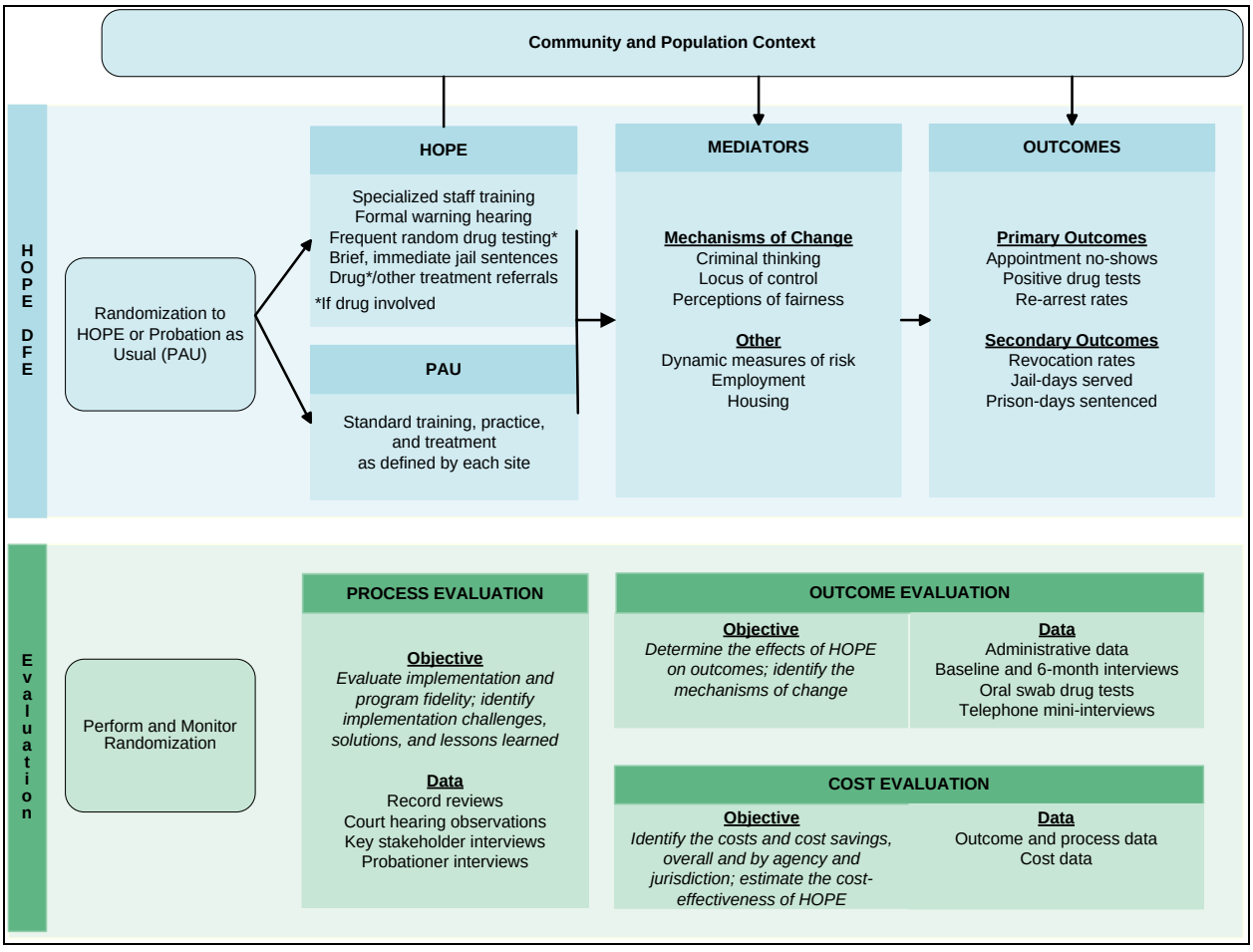
1. Process
    1.1. What was the structural context for the implementation of HOPE in the four sites?
    1.2. Was HOPE implemented with fidelity in the four sites?
    1.3. What lessons were learned for implementation success, replicability, and sustainability?
    1.4. How do intensive drug treatment services offered with the HOPE programs compare with the principles of effective offender intervention?
    1.5. What were the communication pathways among HOPE stakeholders and did these vary from site to site?
    1.6. How did HOPE probationers view their supervision experiences?
2. Outcome
    2.1. Does HOPE participation improve compliance with conditions of supervision and reduce violations?
    2.2. Does HOPE participation reduce recidivism, measured by arrest, conviction, and probation revocation?
    2.3. What is the impact of HOPE on jail days served and prison days sentenced?
    2.4. What is the impact of HOPE on drug use?
    2.5. Does HOPE participation change potential mediators including dynamic recidivism risk factors such as employment and housing stability?

2.6. Does HOPE participation change attitudes that are potential mediators, including participants' criminal thinking/attitudes, perceptions of locus of control, and perceptions of the criminal justice system fairness/legitimacy?

3. Cost
   3.1. What is the cost of starting and implementing HOPE?
   3.2. What are the costs and (any) savings and how are these distributed among the agencies (level of government) participating in HOPE?
   3.3. Is HOPE cost effective?

The DFE and evaluation design are shown in ***Exhibit 1-2***, which shows random assignment to HOPE or probation as usual (PAU); mediators including mechanisms of change (criminal thinking, locus of control, and perceptions of fairness), dynamic risk factors, employment stability, and housing stability; and the primary and secondary study outcomes (appointment no-shows, positive drug tests, re-arrest rates, revocation rates, jail days served, and prison days sentenced). Evaluation components, objectives, and data sources are also shown.

Exhibit 1-2. HOPE DFE model and evaluation design



The evaluation had the following components:

1. ***Eligibility determination and random assignment to HOPE or PAU.*** The study team worked closely with each site to identify who would comprise the local target population; to establish how, when, where, and by whom they would be identified; and to implement an appropriate point and method for randomizing HOPE-eligible probationers to either HOPE or PAU. (Specific procedures for each site are detailed below in the methods section.) The original plan was to randomly assign 400 individuals in each of the four sites—yielding a total study population of 800 HOPE and 800 PAU probationers. Although the enrollment period for the study was extended several times, final enrollment was somewhat less: 1504—743 HOPE and 761 PAU.

- ***Process evaluation/fidelity assessment.*** This component documented the extent to which each program conformed to the HOPE model; documented the barriers, challenges, facilitators, and lessons learned during implementation to fill gaps in the knowledge base as to what is required to set up a HOPE program; and provided evidence as to the generalizability and sustainability of HOPE programs. This component also assessed implementation fidelity of evidence-based drug treatment programs used by the HOPE programs. Implementation and process measures were collected through ***stakeholder interviews; observation of initial warning hearings and court appearances; and review of court, probation, and HOPE project records***. In addition, detailed data were provided to the evaluation team by the program coordinators employed by the sites and overseen by the training and technical assistance providers led by Angela Hawken at Pepperdine University.
- ***Outcome evaluation.*** The outcome study assessed whether HOPE participation improves appointment compliance, drug test results, rearrest rates, revocation rates, jail days served, and prison days sentenced. In addition, the evaluation examined whether HOPE participation changes potential mediators including criminal thinking/attitudes, perceptions of control and justice system fairness and legitimacy, dynamic recidivism risk factors, and employment and housing stability. For the outcome study, ***administrative data*** for all four sites provide information on appointment compliance, drug test results, re-arrests, violations, revocations, and jail and prison days for all HOPE and PAU evaluation participants. We obtained data from local offices (e.g., probation offices, jails, courts), and state agencies (e.g., departments of corrections and probation and parole, state police).[5] ***Interview data*** were collected from evaluation participants at evaluation enrollment (prior to random assignment) and 6 and 12 months post-enrollment on measures to facilitate understanding of the nature of individual change associated with HOPE participation. A random sample of study participants who consented and completed baseline interviews were also asked to participate in a twice-weekly series of mini-interviews (Telephone-Audio Computer Assisted Self interviews or T-ACASI technology). Random ***oral swab drug tests*** were collected on a subsample of HOPE and PAU evaluation participants in conjunction with the follow-up interviews to provide a common measure of current drug use. When permitted, individuals received modest compensation for completing interviews. (See methods section for details.)

[5] Despite repeated requests we were unable to obtain arrest data from the FBI National Crime Information Center (NCIC).

- ***Cost evaluation.*** The cost evaluation combined data from the process and outcome studies and cost estimates from the local sites and the literature to identify costs and costs savings overall and by jurisdiction. In addition, the cost evaluation assessed the cost-effectiveness of the HOPE program model.

2. Evaluation Design and Methods

Planning for the evaluation occurred over a 9-month period and involved site visits and discussions with stakeholders in each of the sites to establish HOPE program eligibility criteria and random assignment procedures that were appropriate to each site, as well as to negotiate any compensation that was to be offered to individuals who agreed to participate in the evaluation. During this time, the study team also developed, programmed, and tested baseline and follow-up data collection instruments; designed the T-ACASI study and instrumentation; and developed protocols and semi-structured interviews for the process study. Discussions were also initiated with agencies in each site in anticipation of negotiating data use agreements for obtaining administrative data in support of the outcome evaluation. Preliminary planning for the cost study also was initiated.

RESEARCH COORDINATORS

Local research coordinators in the four sites solicited participation in the interview component of the study; administered consent and the audio computer assisted interview; observed warning and violation hearings; assisted with administrative data collection; and served as evaluation liaisons to the sites.

During this planning period, the evaluation team also identified and hired research coordinators in the four sites. The research coordinators were in space provided by the local probation offices in three of the four sites (Massachusetts, Texas, Oregon). Because suitable space was not available in Arkansas, the evaluation rented an office across the street from the court house. The role of the research coordinators was to introduce the study to each HOPE-eligible probationer (prior to random assignment) and solicit his or her participation in the interview portion of the study. The research coordinators also administered consent procedures and the audio-computer-assisted-self-interview (ACASI) instruments and oral swab drug tests during follow-up interviews. The research coordinators also served as local liaisons to the evaluation and assisted with administrative data collection in some of the sites.

2.1. Participant Selection and Random Assignment

Eligibility for HOPE was determined by the sites (e.g., HOPE judge, project coordinator,[6] HOPE probation officer, or other probation staff). ***Assignment to the HOPE program was within the discretion of the HOPE judges, and participation in HOPE was not voluntary.*** All eligible probationers were included in the study whether they agreed to participate in the interview portion of the evaluation.

The overall assumptions for HOPE eligibility that were applicable in all sites were:

1. ***High risk.*** Although HOPE was originally proposed for high-risk probationers, medium- or moderate-risk probationers were included in Arkansas and Massachusetts, medium-risk

[6] A HOPE project coordinator or HOPE PC was hired in each site by the agency using grant funds provided through their BJA program grants. The project coordinator oversaw implementation of the HOPE program and worked closely with the HOPE technical assistance provider that was hired by BJA to assist the sites in program startup and implementation. The TA providers worked from Pepperdine University under the direction of Angela Hawken.

cases with violations were included in Oregon and Texas, and low-risk cases with violations were included in Arkansas. The evaluation design recommended a uniform approach to risk determination. Three sites (Arkansas, Massachusetts, and Texas) used the Ohio Risk Assessment System (ORAS) for Community Corrections (or local variant). The Level of Service Case Management Inventory for Community Corrections or the LS/CMI was used in Oregon.

2. ***Substance use.*** Although primarily designed for substance-involved probationers (hence, the focus on drug testing), to replicate the Hawaii demonstration, NIJ and BJA requested that, nondrug-involved individuals who were otherwise eligible for HOPE should be included in the demonstration.
3. ***Time parameters.*** HOPE eligibility was limited to individuals who had 1 year or more remaining on their probation sentences. Individuals newly sentenced to probation were the primary target, but individuals who had served less than 6 months of their probation sentence and had a violation were also eligible for HOPE supervision.
4. ***Exclusion criteria.*** Juveniles, non-English-speaking persons, out-of-county or intrastate transfers, interstate compact, and probationers assigned to some special caseloads were excluded. Exclusions for special caseloads were site dependent. In Massachusetts, these were administrative supervision, pretrial, drug court, ICE custody, and joint probation and parole supervision. In Oregon, these were drug court, driving under the influence, DUI court, ICE custody, mental health court, sex offenders, and mental health caseload, as well as case bank or bench probation. In Texas, these included high-risk/gang, sex offender, and mentally impaired caseloads, as well as parole board cases.

DFE Enrollment Sources and Processes

HOPE-eligible cases were identified from the "flow" of new probationers and the "stock" of existing cases. New probation cases or "flow cases" were identified at the time of sentencing. There were minor variations across the sites; nevertheless, the basic procedures for identifying HOPE-eligible cases and referring these cases for random assignment were similar across sites. Site-specific details are provided later in this section.

Stock cases were individuals who had been on probation for a short period and had recently had one or more probation violations, often as the result of a failed drug test, which resulted in a change in risk status from low (or medium) to high risk. These cases were limited to individuals who had been on probation less than 6 months and who had at least 1 year of supervision remaining. These cases were identified at revocation hearings.

A third source of HOPE-eligible cases was used at the beginning of the study in two sites (Arkansas and Texas). In these sites, the HOPE project coordinator worked with the probation office to generate a list of

PARTICIPANT SELECTION

HOPE eligible probationers were identified from the flow" of new probation cases and the stock of existing probationers with recent violations. Once an individual was randomized to either HOPE or PAU he/she remained in that status throughout the study.

individuals sentenced to probation within 6 months prior to the start of the DFE. The HOPE project coordinator worked with the HOPE probation officers reviewing the list to identify HOPE-eligible probationers, who were asked to report to the probation office for potential reassignment.

After the judge sentenced the offender to probation (or an individual was otherwise identified as a potential HOPE-eligible as noted above), the probation office administrator collected basic information from the probationer and referred the probationer to either a HOPE probation officer or to intake/assessment personnel who completed the risk screener or assessment to identify high (or medium) risk, determined whether any exclusion criteria applied, and completed a study referral slip that was provided by the evaluation. These referral slips were duplicate no-carbon-required (NCR) forms that included a preprinted sequential Study ID Number;[7] blanks for entry of date, name, date of birth, and state ID number; and, on one copy of each duplicate a scratch-off label that obscured the randomly assigned treatment condition. An example of the form is shown in ***Exhibit 2-1***.

The probationer with the referral slip was then sent to the evaluation research coordinator, who introduced the interview portion of the study. These procedures were established so that HOPE-eligible individuals could be approached by the research coordinators for completion of a baseline interview prior to random assignment to HOPE or PAU.[8]

The research coordinator described the purpose of the interviews, what would be required (initial interview that s/he would complete and that would take 30 minutes, opportunity to complete additional interviews in 6 and 12 months), what they would receive ($5 equivalent, for example a McDonald's gift card), and that participation (or not) would have no influence on their supervision. If the individual agreed to participate, the research coordinator reviewed the consent form and obtained consent (witnessed not signed), showed the respondent how to use the computer, and then left the individual to complete the ACASI interview. Once the interview was completed by those who had agreed, the research coordinator introduced the T-ACASI substudy for a random subset of participants. This involved agreement to call into an 800-number twice a week and answer a small number of questions that comprised some of the attitudinal scales included in the full ACASI interview. If the individual agreed to this study, s/he was provided with instructions on when and how to call the line.

Once the interview was completed (or if the individual refused), s/he was sent back to probation for further processing. If the individual refused to participate in the interview, the individual was still subjected to random assignment to either HOPE or PAU. The evaluation obtained administrative data to the extent that local policy provided access for all randomly assigned HOPE-eligible probationers, which was true in all sites.

Once the probationer returned to the probation office, the intake supervisor or other probation personnel retrieved the referral slip from the probationer and scratched off the label that had obscured the study assignment. HOPE was printed beneath the label if the assignment was to HOPE supervision;

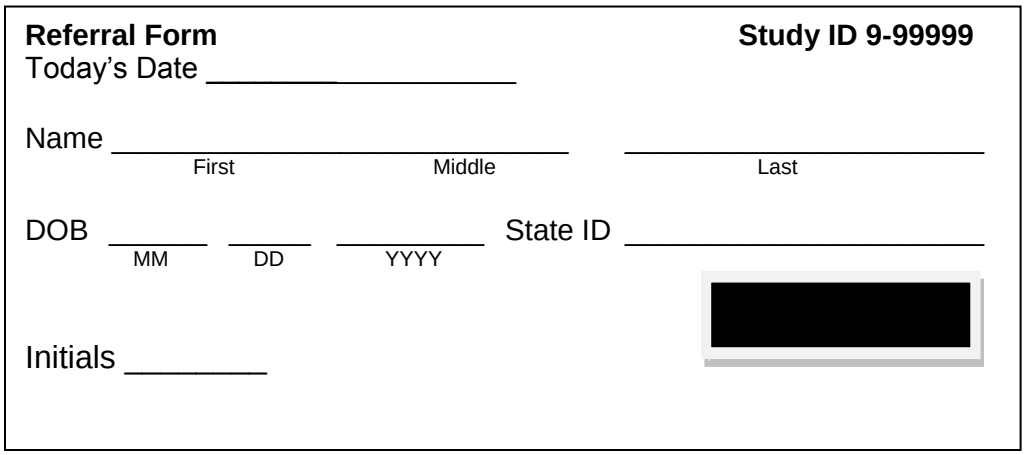
[7] The forms were preprinted with sequential study id numbers to maintain control of the randomization process. Site personnel completed the form for each eligible probationer and sent the form to the research coordinator prior to knowing the assignment of the individual which occurred only after the individual returned to probation after being enrolled in the study by the research coordinator. The evaluation team maintained the crosswalk of study id-study assignment matches on their servers on RTI's campus in North Carolina.

[8] In almost all cases, the interview, if the probationer consented, was completed before he or she knew his or her assignment; in a few cases, the interview was conducted within a few days of assignment.

nothing was printed beneath the label if the assignment was PAU. If the assignment was to HOPE, the probationer was referred immediately to the HOPE probation officer for introduction to HOPE probation; otherwise, the probationer was told to report to a regular probation officer.

Once an individual was randomly assigned to either HOPE or PAU, he or she remained in that status throughout the evaluation period. In other words, a study participant could not be randomly assigned if they once again became HOPE eligible, for example, because of being sentenced to a new probation term. Operationally, this meant that in a few cases individuals were randomized twice but the second random assignment was subsequently voided.

Exhibit 2-1. Example of referral form including covered HOPE/PAU indicator



Enrollment Timing

HOPE program startup and enrollment in the DFE began in August 2012 in Arkansas, Oregon, and Texas and in October 2012 in Massachusetts. The original study design anticipated randomly assigning 400 eligible probationers in each site to HOPE or PAU over a 9-month enrollment period, suggesting that enrollment would be complete in three sites by May 2013 and in the fourth site by June 2013. In fact, enrollment proved much slower than originally anticipated and continued until December 31, 2013 in Oregon; July 31, 2014 in Massachusetts; and September 30, 2014 in Arkansas and Texas. Even with the extended enrollment period, enrollment fell somewhat short of 400 eligible probationers in each site.

Site-Specific Impact Designs

The following sections describe procedures specific to each of the four sites. These procedures were developed by key stakeholders in each site in consultation with the evaluation team. Key participants in the process were the HOPE judge, HOPE probation officers, probation office administrator, HOPE project coordinator, and HOPE research coordinator.

Saline County (Benton), Arkansas

The Saline County probation office has an active caseload of between 500 and 700 probationers. During the DFE, the office had two HOPE probation officers and 6 other probation officers. HOPE-eligible cases were identified from two "flows" (new probation cases and probation revocation arraignments) and one "stock" (HOPE-eligible individuals who were on probation at the time of study start up and had been on probation for less than 6 months). New probation cases and probation revocation arraignment

cases (i.e., revoked probationer to remain on probation) were identified at sentencing. The probation office administrator then collected basic information on the case; a HOPE probation officer then completed the ORAS screener to identify high and medium risk (as opposed to low risk), determined whether any exclusion criteria applied, and completed the referral slip provided by the evaluation. The HOPE probation officer then referred eligible newly sentenced probation cases to the research coordinator. The referral slip was given to the research coordinator (either by the HOPE probation officer or the probationer). The research coordinator provided information on the study to the probationer and followed the procedures outlined above.

Stock cases were identified at the beginning of the evaluation by the HOPE project coordinator, who worked with the probation office to generate a list of probationers who were sentenced to probation during the 6 months prior to study start up (circa February through July 2012). The list was reviewed by the HOPE project coordinator and probation officers to identify HOPE-eligible cases, who were then asked to report to the probation office for potential reassignment. When the individual came in, the HOPE project coordinator or other probation personnel completed the referral slip and referred the probationer to the research coordinator. Procedures were then as described above.

Additional stock cases (implemented by Saline County during 1st week of intake) were also identified among probationers who had been on probation for 6 months or less, and were originally assessed as minimum risk. These individuals were re-assessed by a HOPE probation officer if within the previous 3 months they: had a new arrest, had a positive drug screen, had failed to pay ordered fines, had missed a (or multiple) scheduled appointment(s) with his or her probation officer, or was not attending required treatment. The HOPE probation officer contacted any probationer meeting one or more of these criteria and ordered them to report to the HOPE office for reassessment. If the probationer continued to be ranked as minimum risk, they continued to see their current probation officer and remained on PAU. If the probationer was reassessed as medium or high risk, the HOPE probation officer referred them to the HOPE project coordinator or other probation personnel who completed the referral slip and referred the probationer to the research coordinator. Procedures were then as described above.

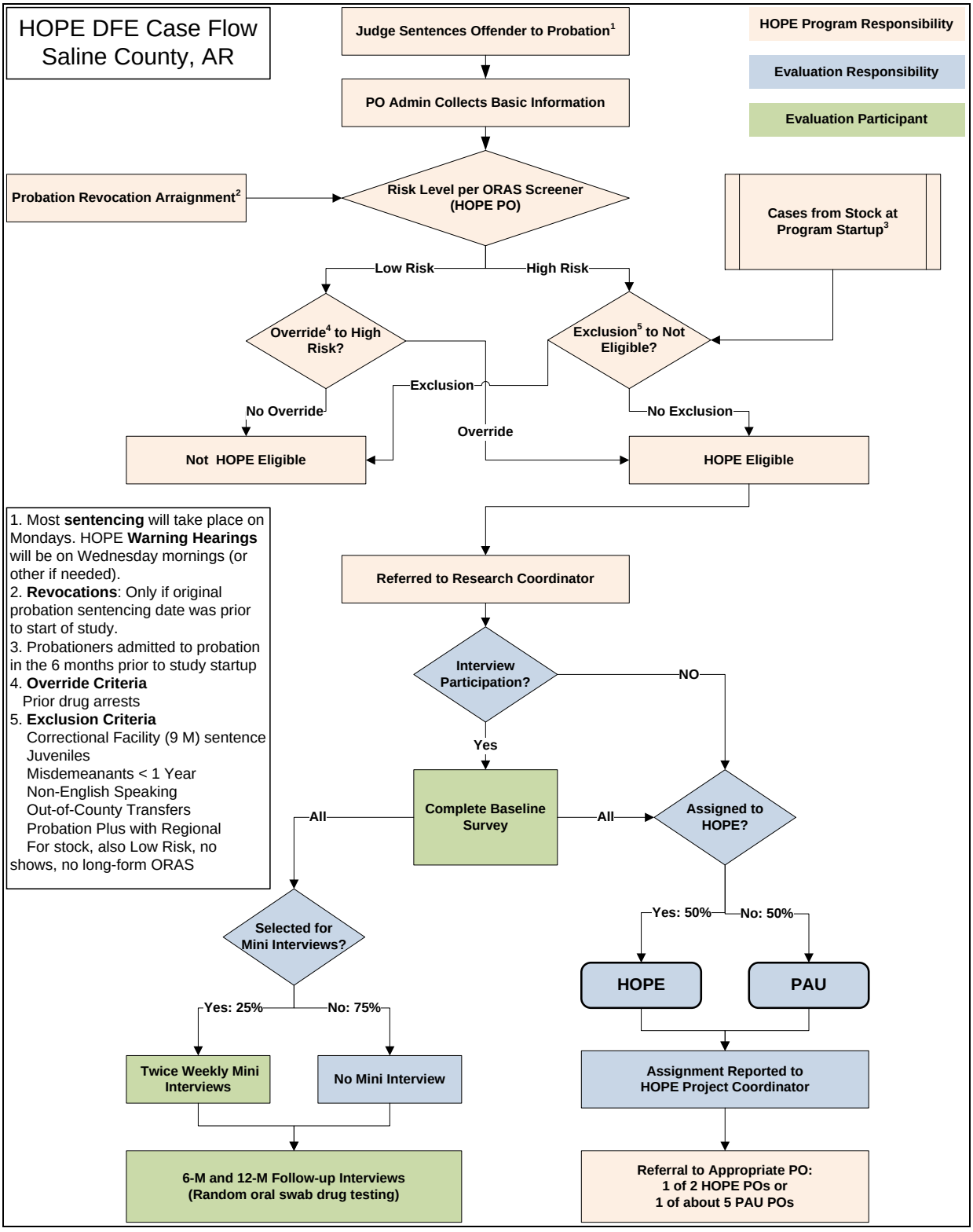
Exclusion criteria for the Saline County site were:

- Juveniles
- Misdemeanants with less than 1-year sentences
- Non-English speaking
- Probation Plus with 9-month sentence to Regional Correctional Facility
- Some out-of-county transfers (not including those living near county line)

In addition, other exclusions from the revocation arraignment pool were no shows, those screened as low risk, and those for whom an ORAS long form was not completed.

The HOPE DFE case flow for Saline County, Arkansas, is shown in ***Exhibit 2-2.***

Exhibit 2-2. Case flow diagram for Saline County, AR



Essex County (Salem), Massachusetts

Massachusetts implemented HOPE in the Salem District Court and the Essex County Superior Court in Salem. These two courts are co-located in Salem but have separate court operations and are served by separate probation offices. District Court cases are those sentenced to straight probation or sentenced to jail plus a probation term. Superior Court cases are individuals being released from prison to an active probation case load. The Essex County site was unique among the four DFE sites in not having dedicated HOPE probation officers. Instead, the HOPE cases were managed by the officers who also had regular caseloads[9].

District Court cases were identified at the time of sentencing to supervised probation and referred to Probation Administrators for risk screening using the ORAS. Superior Court cases being released from Massachusetts' prisons to Essex County probation supervision were assumed to be high risk (by nature of their prison sentences) and were required to report to the Superior Court HOPE judge within 48 hours of prison release. Referral slips were completed by probation for the high-risk cases that were not subject to the exclusion criteria and the HOPE-eligible probationers were then referred to the research coordinator for introduction to the interview portion of the study.

In Massachusetts, ineligibility criteria included those on administrative supervision, pretrial, drug court, ICE custody, and joint probation and parole supervision.

The following probationers were *ineligible* from District Court:

- On administrative supervision
- Pre-trial probation
- Does not live in Court jurisdiction
- Drug Court
- Immigration and Customs Enforcement (ICE) custody
- Interstate Compact Offender Tracking System (ICOTS)
- Not English-language proficient
- Intra-state transfers
- Joint Probation and Parole supervision

The following probationers were *ineligible* from Superior Court:

- Scores "low" risk on ORAS short form (unless there is override to "high")
- On administrative supervision
- Continuance without a finding (CWOFs)[10]/Pre-trial probation
- Does not live in Court jurisdiction
- Drug Court
- ICE custody
- ICOTS
- Not English-language proficient

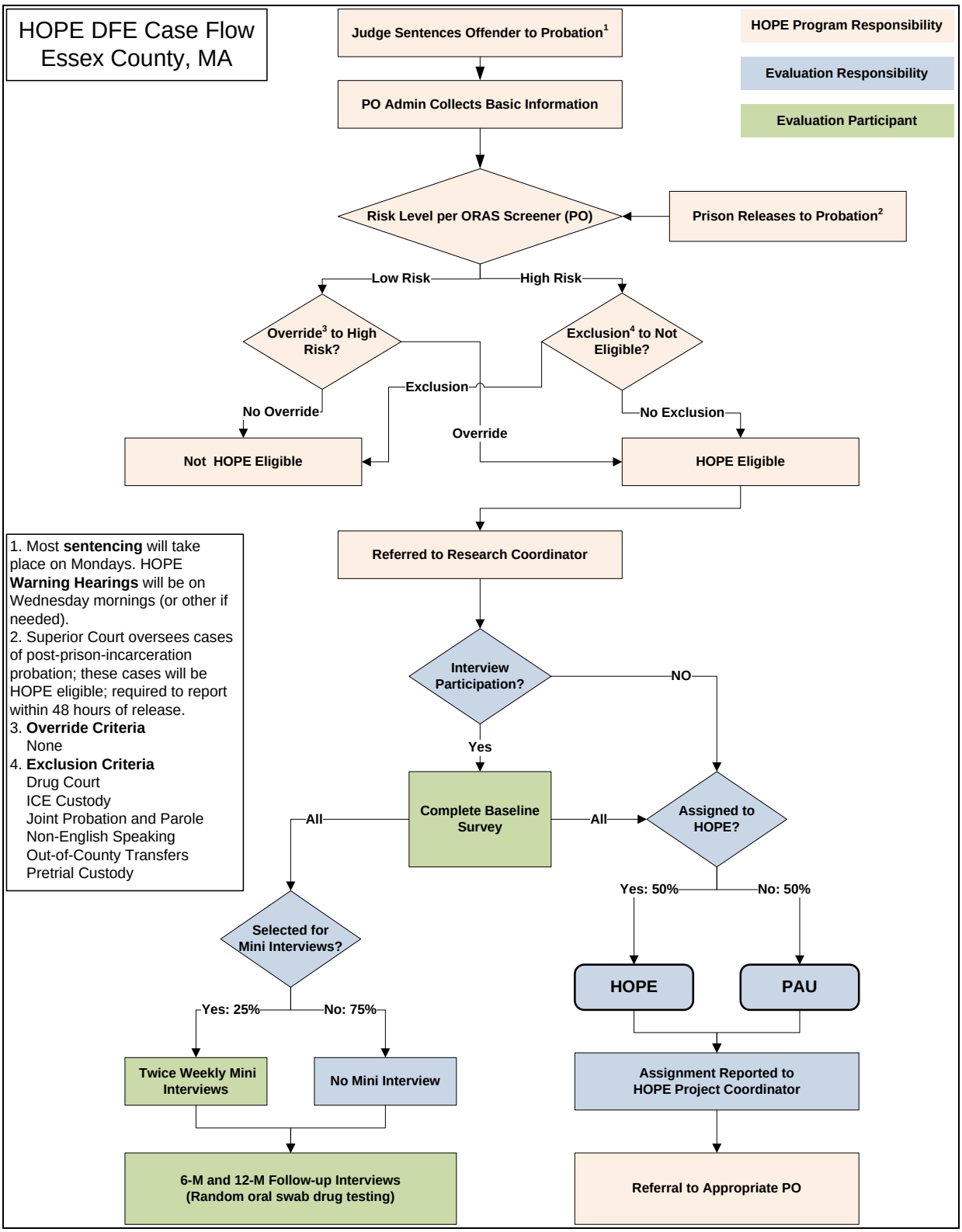
[9] Evaluation PAU cases were not assigned to be supervised by the probation officers who were managing HOPE caseloads. In other words, the officers saw HOPE cases and other individuals who were not part of the DFE.

[10] A continuance without a finding, or CWOF, is a resolution of a criminal complaint via an agreement between the Prosecutor and Defendant that the Prosecutor can demonstrate sufficient facts to the Court that the Defendant engaged in the alleged criminal activity.

- Joint Probation and Parole supervision

The HOPE DFE case flow for Essex County is shown in ***Exhibit 2-3.***

Exhibit 2-3. Case flow diagram for Essex County, MA



Clackamas County (Oregon City), Oregon

Clackamas County identified high-risk probationers as HOPE-eligible from the following flows of probationers:

- New probation cases identified as high risk by the PSC screener at intake
- New probation cases initially identified as medium risk by the PSC screener subsequently assessed as high risk on the LS/CMI at initial assessment
- New probation cases initially assessed as medium risk who are reassessed as high risk on the LS/CMI at reassessment (6 months post initial assessment or sooner)
- Revoked probationers who had spent 60 days in jail and were returned to probation under local control who still have 1 year of probation remaining

For new probation cases, the judge's probation order was sent to Probation intake, where it was reviewed and the OCMS or PSC was generated to identify risk level. The probationer was also ordered to appear at intake within 24 hours or sentencing or release from any jail sentence. For high risk cases, intake staff determined whether any exclusion criteria applied (in some cases, this step was completed the previous day during the initial review). If the case was determined to be HOPE-eligible, intake staff completed the referral slip, and referred the probationer to the research coordinator. The referral slip with the HOPE/PAU designation under the scratch-off label was retained by intake, while the duplicate copy of the slip was given to the research coordinator who entered the information on the form into the computer to initiate a new case.

New probation cases initially screened as medium risk that were subsequently assessed on the initial LS/CMI as high risk began their supervision assigned to a probation officer based on risk and probation officer rotation. Within 30 days of assignment, the probationer met with the probation officer who completed the initial LS/CMI assessment. If the probationer was assessed as high risk on this assessment, s/he was taken by the probation officer to intake for further processing as was described above. The HOPE project coordinator ran periodic reports to check that this referral to intake and to the research coordinators took place.

New probation cases initially screened as medium risk that were reassessed on the LS/CMI as high risk within 6 months (or so) of initial sentencing were processed as described for those who were initially screened as medium risk but were assessed as high risk during their initial assessment.

Probation cases revoked to 60 days in jail and released to local control were assumed to be high risk (some exclusions applied—e.g., child support cases). For these cases, the local control administrator received notification that a local control case was to be released from jail. These cases were directed to report to community corrections within 24 hours of release. HOPE-eligible local control probationers were directed by the front desk to report to intake, where staff completed the referral slip and referred the individual to the research coordinator, where the case was processed as described above.

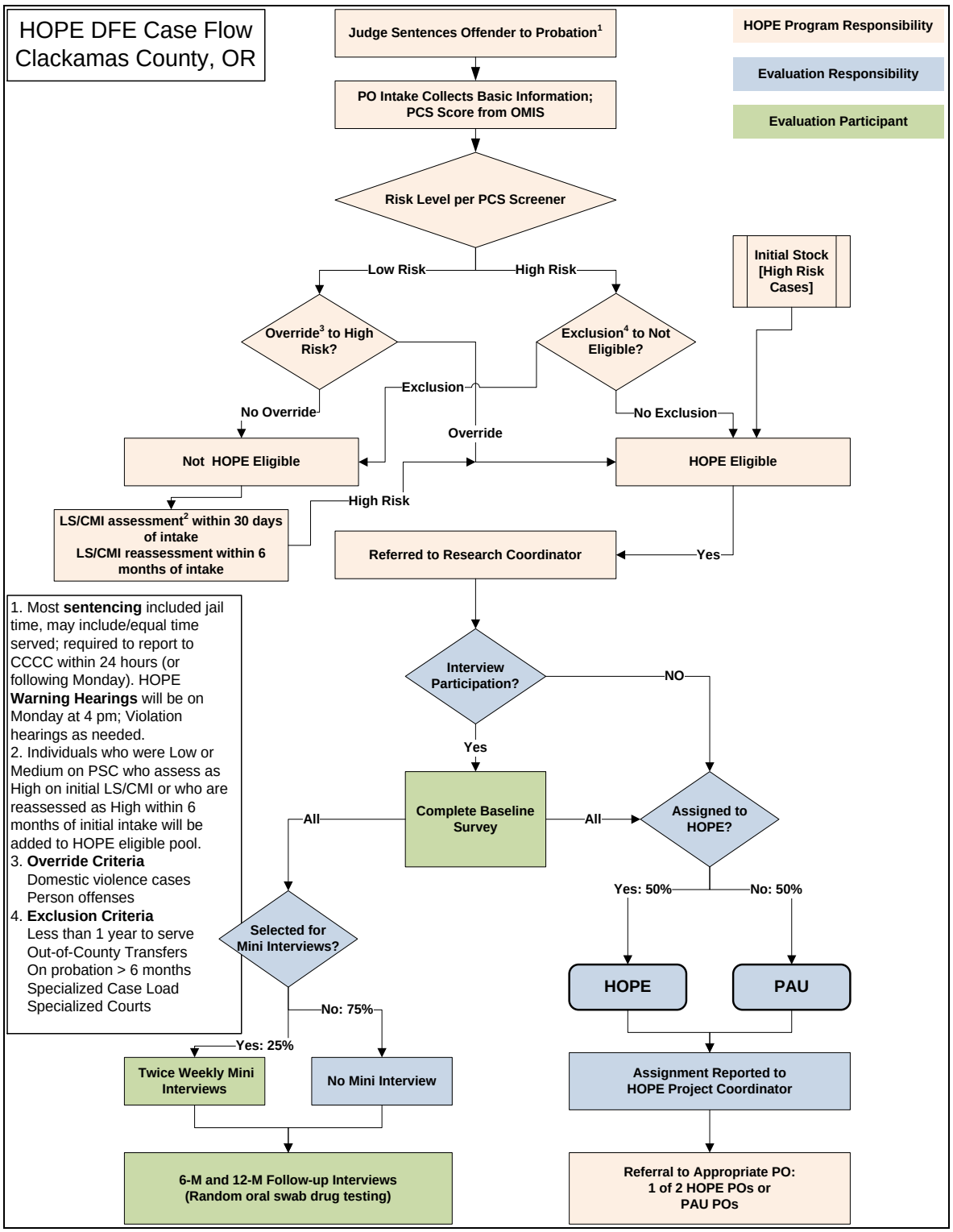
The following exclusion criteria were used in Clackamas County:

- Bank or bench probation
- Less than 1 year to serve
- Out-of-county transfers
- Had served more than 6 months of the current probation term
- Specialized caseloads (e.g., sex offenders, and mental health caseloads)

- Specialized courts (e.g., drug court, influence, DUI court, ICE custody, mental health court)

The HOPE DFE case flow for Clackamas County, Oregon, is shown in ***Exhibit 2-4.***

Exhibit 2-4. Case flow diagram for Clackamas County, OR



Tarrant County (Ft. Worth), Texas

Tarrant County (Fort Worth), which already operated a SWIFT court for regular high-risk probationers, focused the HOPE program on their State Jail Felon (SJF) population. SJFs have been convicted of Level 4 felonies—mostly property and drug offenses, no violent offenses. These individuals are on probation in lieu of serving a state jail sentence (e.g., 2 years in state jail or 5 years of probation); state jail time is straight time (no good time is provided, i.e. individuals cannot earn sentence reductions through, e.g., days with good behavior or completion of programs). Texas reported that up to 90% of SJFs are revoked because of a high level of violations. They are eligible for up to 6 months of sanction time (the same as all felony classes) and are "hard to manage" because of limited options for sanctioning.

There were two "flows" (new high/very high risk probation cases and stock populations of medium risk probationers who incurred a violation) and one "stock" (high/very high risk HOPE-eligible individuals who were on probation at the time of study start up and had been on probation for less than 6 months) from which HOPE-eligible cases were identified.

New probation cases were sent to intake for processing (photo, etc.) and to get their initial field intake appointment (based on zip code). They were then sent to assessment to make appointment for assessments (ORAS, TCUDS, MH screen) within the next couple of days. At the assessment appointment, the assessment officer determined if the case was HOPE eligible—i.e., assessed high/very high on ORAS, English speaking, lives in Tarrant County, had at least 30 days of probation jail time remaining, not an out-of-county transfer, and not eligible for a specialized case load (e.g., substance abuse felony punishment facility (SAFPF), Intensive Day Treatment (IDT) program, sex offender, seriously mentally ill (MI)). Sex offenders who had failed to register and who had no (other) sex offender conditions were eligible. HOPE-eligible cases were sent to the HOPE project coordinator who completed the referral slip and referred the cases to the research coordinator, who entered information from the referral slip into the computer and introduced the interview portion of the study.

Medium risk probationers with new violation were identified by the HOPE project coordinator, who worked with the probation office to generate monthly a list of medium-risk SJFs who had received a probation violation in the previous period (e.g., month) and had been on probation less than 6 months. This query was run on the first day of each month to identify medium-risk offenders who had a violation in the previous month. The HOPE project coordinator then reviewed the list to determine eligibility. The HOPE-eligible cases were asked to report to the probation office and meet with the HOPE project coordinator for potential reassignment. The HOPE project coordinator completed the referral slip, and referred the probationers to the research coordinator, who entered information from the referral slip into the computer and introduced the interview portion of the study.

During DFE start-up, HOPE-eligible cases were also identified from the stock of SJF probationers who had been sentenced to probation in the 6 months prior to study start up (circa February through July 2012). The HOPE project coordinator worked with the probation office to generate a list and reviewed the list with the HOPE probation officers to identify HOPE-eligible cases that were asked to report to the probation office for potential reassignment. Procedures were then as described above.

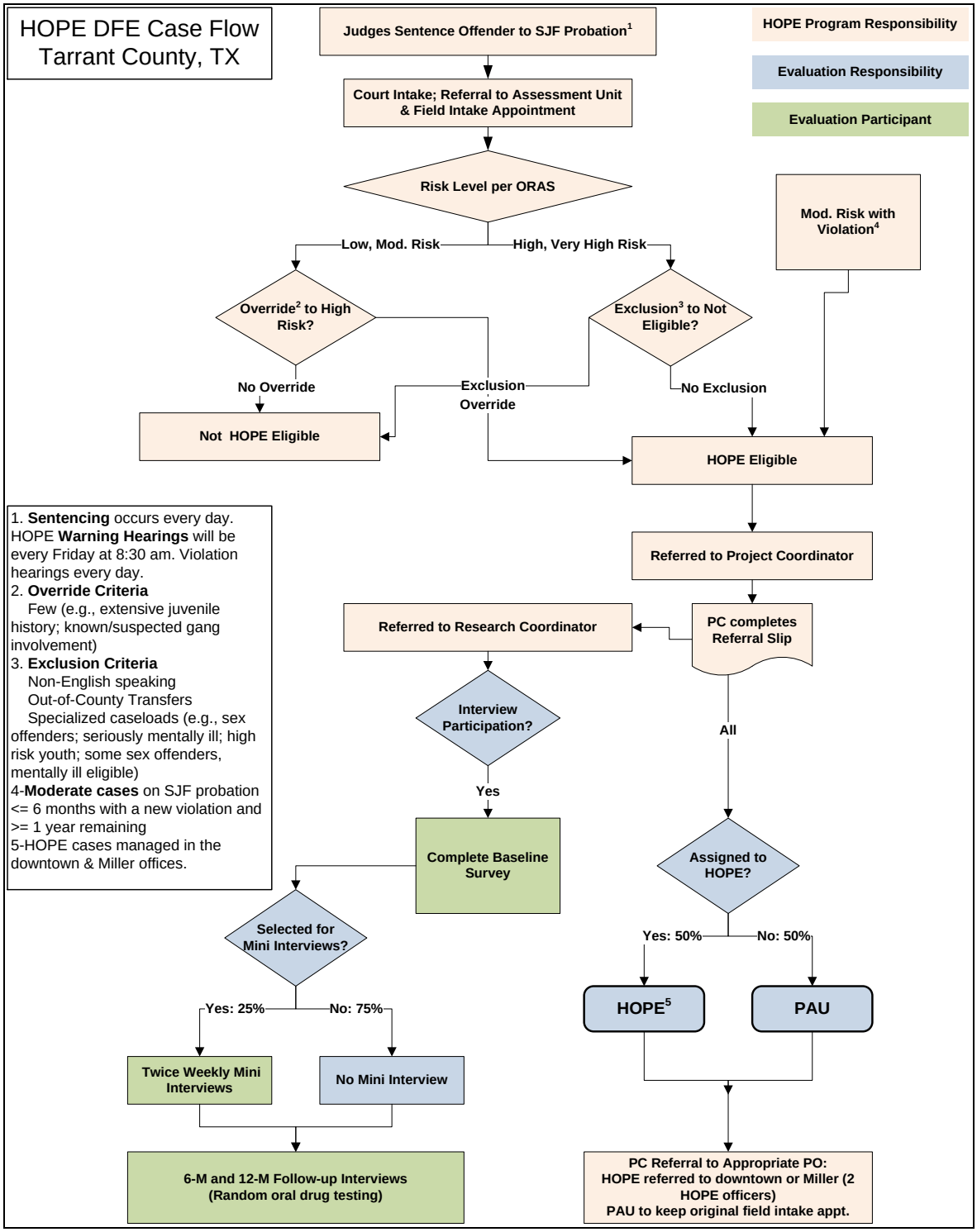
In Texas, the following caseloads were not eligible for HOPE supervision:

- High-risk/gang
- Parole board cases

- Sex offender and mentally impaired

The HOPE DFE case flow for Tarrant County, Texas, is shown in ***Exhibit 2-5.***

Exhibit 2-5. Case flow diagram for Tarrant County, TX

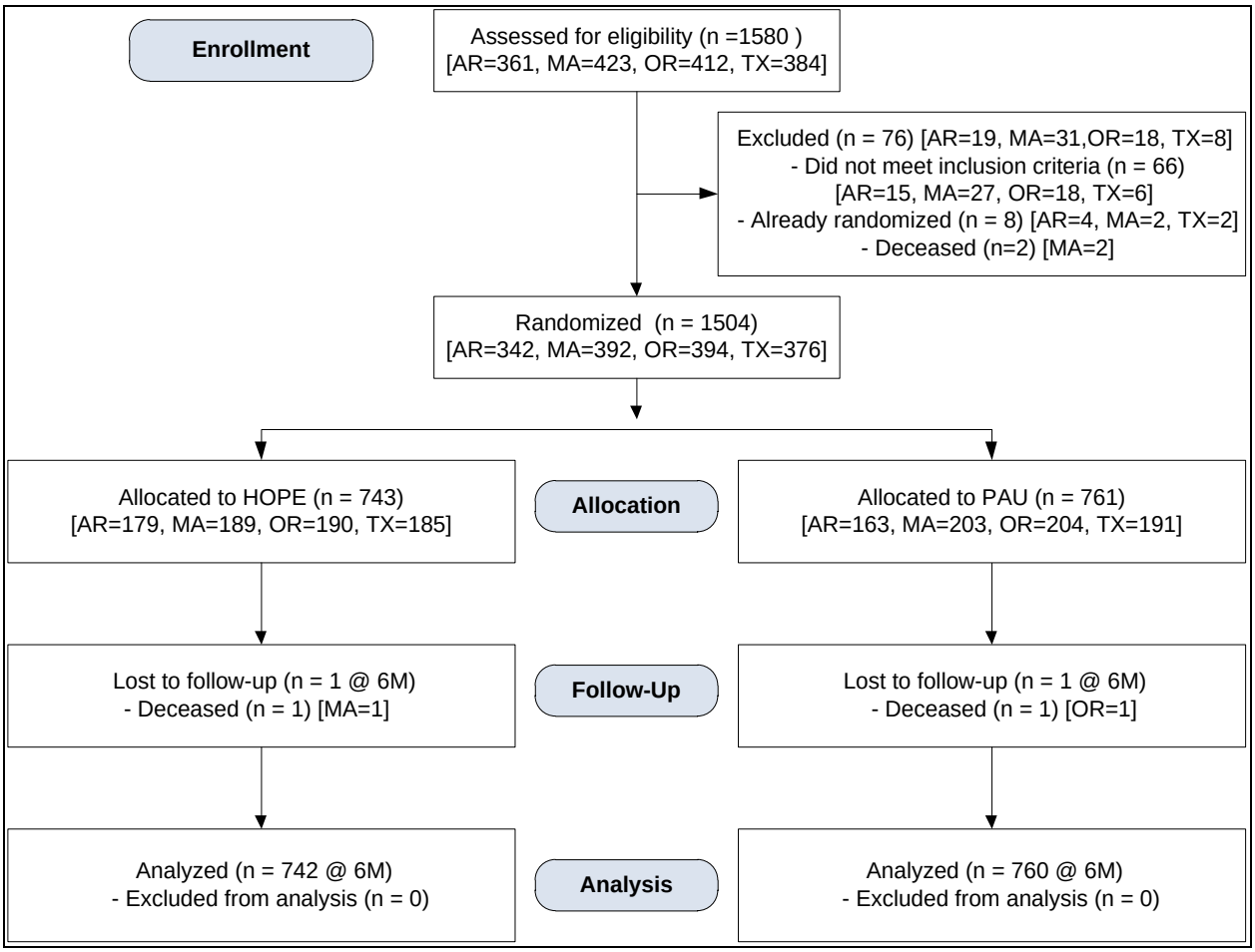


RCT Registration

The evaluation used the Consolidated Standards of Reporting Trials (CONSORT) Statement (Schulz, Altman, and Moher, 2010), developed for medical trials, to develop, monitor, and describe the HOPE DFE. ***Exhibit 2-6*** shows a CONSORT flow diagram for the site-level HOPE DFE RCT. RTI registered the Multisite HOPE RCT with ClinicalTrials.gov (NCT01670708) and with the Cambridge Criminology Registry of Experiments in Correctional Strategy and Tactics (REX-COST).

The HOPE DFE was registered with ClinicalTrials.gov as the Multisite Evaluation of the Honest Opportunity with Enforcement Demonstration Field Experiment (HOPE DFE Multisite Evaluation, registration number NCT01670708) in August 2012. Final CONSORT (Consolidated Standards of Reporting Trials) results for the study are shown in ***Exhibit 2-6*** (Schulz, Altman, and Moher, 2010), overall and by site. A total of 1,580 individuals were assessed for eligibility, of which 76 were subsequently excluded primarily because they were determined to be ineligible for HOPE (n = 66) after they were initially screened into the DFE. The 1,504 remaining individuals were HOPE (n = 743) or PAU (n = 761). Two subjects were known to have died after random assignment but before time for their 6-month interview (13 participants died during the study). Final numbers of individuals for analysis were 342 (Arkansas), 392 (Massachusetts), 394 (Oregon), and 376 (Texas).

Exhibit 2-6. CONSORT diagram for the HOPE DFE RCT



2.2. Process Evaluation: Implementation and Fidelity

Program effectiveness is a product of the measured effect size plus the quality and fidelity of implementation (Tucker and Roth, 2006). Effective programs are those that adopt beneficial program models and implement them well. It is insufficient to do one or the other. A structured process evaluation is key to developing a full understanding of program operations and the inferences to be derived from outcome evaluations. The process evaluation, and especially the implementation fidelity analysis, was designed to inform whether the program was delivered as intended (fidelity to the model) and to document challenges and issues surrounding implementation.

PROCESS EVALUATION

The process evaluation evaluated the fidelity of implementation to the intended HOPE model and documented the implementation experience at the four DFE sites.

The process evaluation addressed the following research questions:

1. What was the structural context for the implementation of HOPE in the four sites?
2. Was HOPE implemented with fidelity in the four sites?
3. What lessons were learned for implementation success, replicability, and sustainability?
4. How do intensive drug treatment services offered with the HOPE programs compare with the principles of effective offender intervention?
5. What were the communication pathways among HOPE stakeholders and did these vary from site to site?
6. How did HOPE probationers view their supervision experiences?

Fidelity Analysis

Implementation analysis is a key component to any program evaluation agenda (Hatry, Winnie and Fisk, 1981; Rossi, Lipsey and Freeman, 2003). Effective programs employ specific activities and interventions known to produce desired outcomes (intervention effectiveness), and implement those interventions with high fidelity to design (implementation fidelity) (Fixsen, Naoom, Blase, Friedman and Wallace, 2005). A program may select or design evidence-based interventions but implement them poorly (high intervention effectiveness, low implementation fidelity); conversely, a program may select or design poor interventions, but implement them well (low intervention effectiveness, high implementation fidelity). ***Exhibit 2-7*** below summarizes these possibilities, with the goal of moving programs into Quadrant 1, which is where client outcomes are maximized.

A growing body of literature indicates that social programs that maintain a high degree of fidelity between program design/theory (interventions) and program practice (implementation) show better outcomes than those that do not, with program effects up to three times as large for high-fidelity programs (Andrews and Dowden, 2005; Durlak and DuPre, 2008; Fixsen et al., 2005; Gottfredson and Gottfredson, 2002; Lipsey, 2009; Olds, 2002), and with potentially iatrogenic effects when fidelity is poor (Washington State Institute for Public Policy, 2002).

Exhibit 2-7. Interaction of intervention effectiveness and implementation fidelity

		Implementation Fidelity	
		Low	**High**
Intervention Effectiveness	**Effective**	***Quadrant 2*** Good Intervention Poor Implementation	***Quadrant 1*** **Good Intervention** **Good Implementation**
	Ineffective	***Quadrant 3*** Poor Intervention Poor Implementation	***Quadrant 4*** Poor Intervention Good Implementation

One cannot assume that a program was delivered as intended (Esbensen, Matsuda, Taylor and Peterson, 2011). One must measure it. Deviations from intended program logic are common and can explain the failure of programs that otherwise "should" have worked (Durlak and DuPre, 2008; Dusenbury, Brannigan, Falco and Hansen, 2003; Fagan, 2013). Well-cited examples of "failure by poor implementation" include variations in provider quality during the statewide roll out of the Functional Family Therapy program in Washington state (Washington State Institute for Public Policy, 2002) and the failure to engage qualified nurses in some communities that adopted the Nurse-Family Partnership program (Olds, 2002), resulting in differential outcomes across those communities.

Practice settings such as criminal justice agencies may have legitimate reasons for modifying and adapting programs to their needs, and many intervention models can tolerate some degree of local innovation without a fatal threat to intervention effectiveness, but it is nonetheless important to know how well the program in practice adhered to the program theory (Elliott and Mihalic, 2004; Rossi, Lipsey, and Freeman, 2003). Without this assessment, there is a risk of attributing outcomes (positive, null or iatrogenic) to program processes that were not present, or which were only weakly provided (Fagan, 2013).

The approach employed for assessing implementation fidelity of the HOPE DFE sites involved breaking the program model down into a set of fidelity items that represent essential features of HOPE, and then measuring the extent to which each site enacted those elements in practice. There is general agreement that fidelity assessment is more effective when guided by a checklist or matrix that breaks a given intervention down into its core components, outlining key intended program features and allowing the evaluator to measure whether those features were enacted in practice (Durlak and DuPre, 2008; Esbensen et al., 2011; Fixsen et al., 2005). This checklist allows for a more objective, rigorous and continuous quantitative assessment of implementation fidelity, as opposed to more casual expert observations and judgments about whether a program has done what it is supposed to do.

An initial challenge for the implementation fidelity evaluation was to determine the intended program logic for HOPE – its espoused theory (Argyris, 1985)—against which the actual program operations at the four DFE sites (the theory-in-use) would be gauged. The underlying HOPE model, following from the Hawaii HOPE progenitor, has been explained elsewhere (Hawken and Kleiman, 2009; Zajac et al., 2015). This existing discussion notwithstanding, the operational details of HOPE arguably

have not been as well codified as other more long-standing intervention approaches such as cognitive-behavioral therapy (Van Voorhis and Salisbury, 2013). Moreover, HOPE is in a period of rapid dissemination and uptake across correctional systems nationally, with the sort of innovation and variation that often attends such expansion (Fixsen et al., 2005). Still, a requisite element in designing an implementation fidelity evaluation is establishing exactly what the program is supposed to look like.

We were fortunate that the four program sites in the DFE were selected by the BJA following the evaluation of proposals submitted in response to a solicitation released in 2011–*The Honest Opportunity Probation with Enforcement (HOPE) Demonstration Field Experiment FY 2011 Competitive Grant Announcement*–which invited interested jurisdictions to apply to participate in the DFE as HOPE replication sites. This solicitation outlined key programmatic components that selected sites were required to enact in developing and implementing their HOPE programs. We have taken these programmatic components as the *espoused theory* of HOPE for the purposes of the implementation fidelity analysis within this DFE.

THE HOPE MODEL

Key programmatic elements of HOPE were described in the BJA solicitation for sites to replicate the HOPE model. These eleven elements provide a model for what HOPE was to look like in the DFE replications.

From a content analysis of the solicitation, we identified eleven items that we took to be key markers of expected implementation for the DFE sites. These eleven items represent the essential things that the four sites were supposed to do in their day-to-day operation of HOPE and are (1) central to the underlying HOPE model; (2) implicitly set as expectations on the DFE sites through the Bureau of Justice Assistance solicitation that funded the four HOPE DFE sites; and (3) well within the control of the DFE sites. Further, these are measures for which we have enough data to form conclusions about fidelity with respect to the items. These measures also correspond well to the SCF Success Benchmarks, as discussed in Oleson (2016), which are taken from the swiftcertainfair.com website[11].

Item 1 (Leadership): Program leadership/championship is identified in the implementation literature as important to program implementation (Fixsen et al., 2005). The BJA solicitation established the expectation that clear leadership would be built around HOPE at the DFE sites, most likely revolving around the HOPE judge. We assessed leadership through qualitative interviews with HOPE team members during visits to the DFE sites (discussed further below). We probed for evidence that there was consensus around leadership of HOPE at that site. The leader might be HOPE Judge (as is specified in the HOPE model per the BJA solicitation) or some other team member. Leadership could also shift over time, but the expectation was that some leadership could be documented.

Item 2 (Probationers High Risk): All probationers selected for HOPE under this DFE were supposed to be moderate to high risk using a standardized offender risk assessment tool. This tool was the Ohio

[11] As noted earlier, Judge Alm notes that the Hawaii HOPE model presumed an RNR/MI approach to probation that was the presumed model of probation for HOPE prior to implementing the SCF components. These elements were not discussed in the original Hawaii HOPE evaluation reporting nor were they listed as requirements (either existing or to be implemented) for probation agencies seeking grant funds to implement and test HOPE as part of the DFE.

Risk Assessment System (ORAS) in most cases (Latessa, Smith, Lemke, Makarios and Lowenkamp, 2009). This item assesses the percentage of HOPE probationers who were moderate to high risk.

Item 3 (Warning Hearing Compliance): The formal warning hearing is a key component of the HOPE model. The selected sites were expected to follow a brief written model script established by Judge Alm for the Hawaii HOPE program (with modifications allowed to fit the local context). We established a warning hearing checklist (***Exhibit 2-8***) that DFE research coordinators placed at each site by RTI used to assess the degree of congruence between the model script (allowing for local modifications) and a randomly observed sample of each Judge's warning hearings at each site. This checklist broke down the warning hearing into 14 key themes identified in the model script. Given the presumptive ease of following a brief written script, we required that the Judge meet 12 of these 14 themes to be given credit for compliance with the script in each hearing.

Item 4 (Initial Drug Testing Frequency): The BJA solicitation specifies that sites must conduct at least eight random drug tests on each HOPE probationer during their first 2 months in HOPE. This item relates to the centrality of *surveillance* to the HOPE model. We assessed the percentage of HOPE probationers who received at least eight drug tests during the first 2 months.

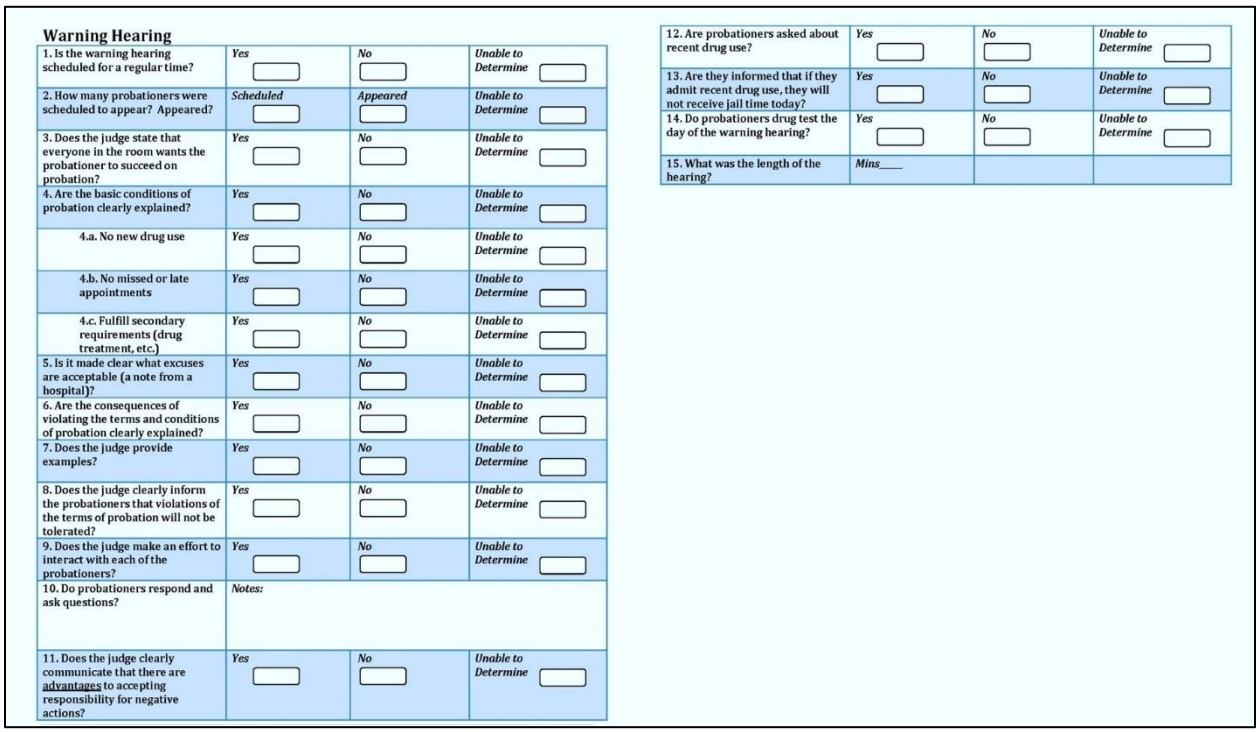
Item 5 (Stepped Down Drug Testing Frequency): The BJA solicitation specifies that sites must conduct at least one random drug test per month on each HOPE probationer after the initial 2-month high intensity testing phase (assuming successful probationer compliance with the first phase of testing). As with Item # 4, we assessed the percentage of HOPE probationers who received at least one drug test per month after the first 2 months.

Item 6 (Exceptions for Missed Drug Testing): Certainty of consequences is one of the central features of the HOPE model. And given the prominence of drug testing in the model, if probationers are permitted to evade drug tests without consequences that would constitute a significant threat to the integrity of HOPE supervision. The BJA solicitation clearly cautions against allowing HOPE probationers to evade drug testing. Thus, we assessed the percentage of missed drug tests that were met with a consequence *unless* a legitimate excuse was provided by the probationer (which for all four sites constituted either being in the hospital or in some form of custody).

Item 7 (Time to Violation Hearing): Swiftness of sanctioning is one of the key precepts of HOPE. This item reflects the time in days between a probation violation and the appearance of the probationer before the judge for a violation hearing. The sites were expected to hold the hearing within 3 days of the violation, per the BJA solicitation. We assessed the percentage of violations that were followed by a hearing within the 3-day window. Certain items were also to be covered in the hearings and a checklist detailing the elements was prepared by the TTA team. The checklist is shown in ***Exhibit 2-9***.

Item 8 (Sanction Type): The BJA solicitation suggests that jail or other confinement is the preferred sanction, as opposed to other consequences such as community service, fines or essays. Thus, we define fidelity here as issuance of a jail stay or other form of confinement (e.g. commitment to a residential treatment facility) in response to a violation. We assessed the percentage of sanctions that consisted of confinement. Note that some violations resulted in multiple types of sanctions, but a case was credited as complying if at least one of those sanctions was confinement.

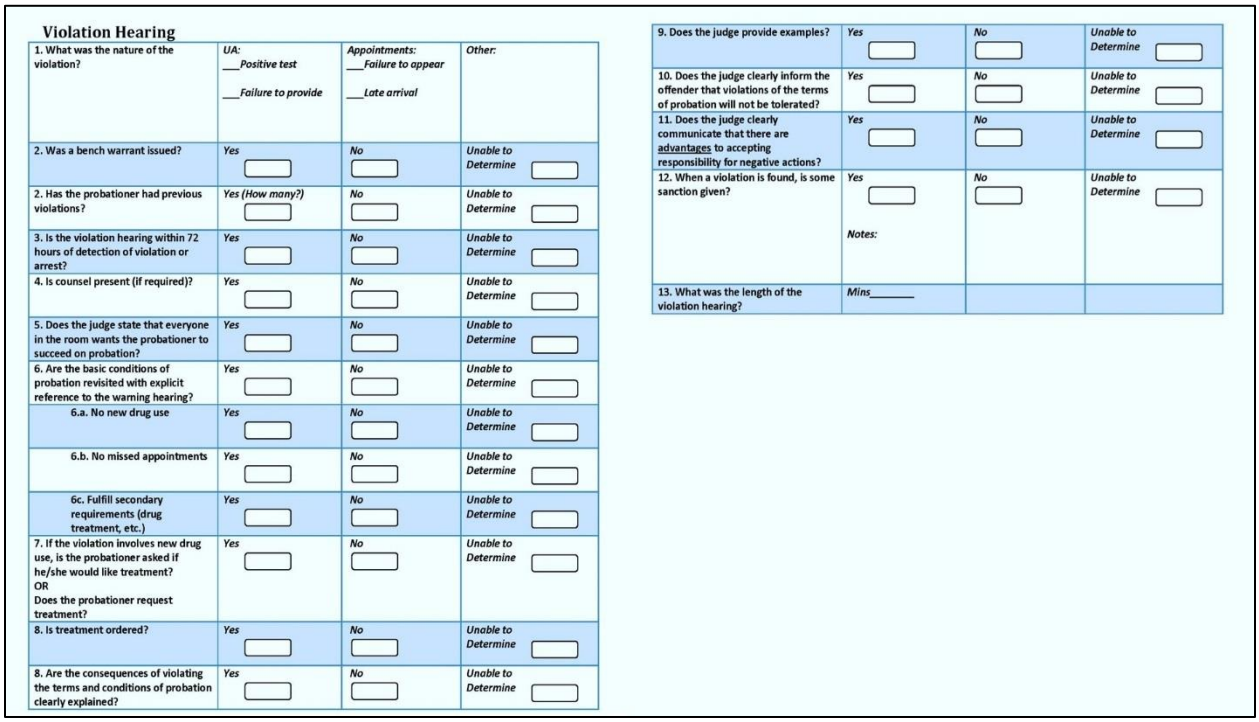
Exhibit 2-8. HOPE DFE warning hearing checklist



Warning Hearing

1. Is the warning hearing scheduled for a regular time?	Yes ☐	No ☐	Unable to Determine ☐
2. How many probationers were scheduled to appear? Appeared?	Scheduled ☐	Appeared ☐	Unable to Determine ☐
3. Does the judge state that everyone in the room wants the probationer to succeed on probation?	Yes ☐	No ☐	Unable to Determine ☐
4. Are the basic conditions of probation clearly explained?	Yes ☐	No ☐	Unable to Determine ☐
4.a. No new drug use	Yes ☐	No ☐	Unable to Determine ☐
4.b. No missed or late appointments	Yes ☐	No ☐	Unable to Determine ☐
4.c. Fulfill secondary requirements (drug treatment, etc.)	Yes ☐	No ☐	Unable to Determine ☐
5. Is it made clear what excuses are acceptable (a note from a hospital)?	Yes ☐	No ☐	Unable to Determine ☐
6. Are the consequences of violating the terms and conditions of probation clearly explained?	Yes ☐	No ☐	Unable to Determine ☐
7. Does the judge provide examples?	Yes ☐	No ☐	Unable to Determine ☐
8. Does the judge clearly inform the probationers that violations of the terms of probation will not be tolerated?	Yes ☐	No ☐	Unable to Determine ☐
9. Does the judge make an effort to interact with each of the probationers?	Yes ☐	No ☐	Unable to Determine ☐
10. Do probationers respond and ask questions?	Notes:		
11. Does the judge clearly communicate that there are advantages to accepting responsibility for negative actions?	Yes ☐	No ☐	Unable to Determine ☐
12. Are probationers asked about recent drug use?	Yes ☐	No ☐	Unable to Determine ☐
13. Are they informed that if they admit recent drug use, they will not receive jail time today?	Yes ☐	No ☐	Unable to Determine ☐
14. Do probationers drug test the day of the warning hearing?	Yes ☐	No ☐	Unable to Determine ☐
15. What was the length of the hearing?	Mins____		

Exhibit 2-9. HOPE DFE violation hearing checklist



Violation Hearing

1. What was the nature of the violation?	UA: ___Positive test ___Failure to provide	Appointments: ___Failure to appear ___Late arrival	Other:
2. Was a bench warrant issued?	Yes ☐	No ☐	Unable to Determine ☐
2. Has the probationer had previous violations?	Yes (How many?) ☐	No ☐	Unable to Determine ☐
3. Is the violation hearing within 72 hours of detection of violation or arrest?	Yes ☐	No ☐	Unable to Determine ☐
4. Is counsel present (if required)?	Yes ☐	No ☐	Unable to Determine ☐
5. Does the judge state that everyone in the room wants the probationer to succeed on probation?	Yes ☐	No ☐	Unable to Determine ☐
6. Are the basic conditions of probation revisited with explicit reference to the warning hearing?	Yes ☐	No ☐	Unable to Determine ☐
6.a. No new drug use	Yes ☐	No ☐	Unable to Determine ☐
6.b. No missed appointments	Yes ☐	No ☐	Unable to Determine ☐
6c. Fulfill secondary requirements (drug treatment, etc.)	Yes ☐	No ☐	Unable to Determine ☐
7. If the violation involves new drug use, is the probationer asked if he/she would like treatment? OR Does the probationer request treatment?	Yes ☐	No ☐	Unable to Determine ☐
8. Is treatment ordered?	Yes ☐	No ☐	Unable to Determine ☐
8. Are the consequences of violating the terms and conditions of probation clearly explained?	Yes ☐	No ☐	Unable to Determine ☐
9. Does the judge provide examples?	Yes ☐	No ☐	Unable to Determine ☐
10. Does the judge clearly inform the offender that violations of the terms of probation will not be tolerated?	Yes ☐	No ☐	Unable to Determine ☐
11. Does the judge clearly communicate that there are advantages to accepting responsibility for negative actions?	Yes ☐	No ☐	Unable to Determine ☐
12. When a violation is found, is some sanction given?	Yes ☐ Notes:	No ☐	Unable to Determine ☐
13. What was the length of the violation hearing?	Mins_____		

Item 9 (Sanction Dosage): The premise of HOPE is that swiftness and certainty are the key components of sanctioning; sanction severity is less important, and indeed, extreme severity is to be

avoided. The BJA solicitation offers no clear guidance on the ideal number of jail days, and this can vary somewhat depending upon history of violations and other factors. Hawken and Kleiman (2009) report that there was considerable variation among the judges in the original Hawaii HOPE evaluation in the number of jail days assigned for a violation, ranging from a few days to as much as 6 weeks. A "typical" sanction though seemed to hover around the 1-week mark. The mean total jail days for violations for a given probationer in the Hawaii study was 19, but again, this could reflect the accumulation of several violations for some probationers, or a single violation for others (potentially indicating that the violation was particularly egregious, or the judge in question followed a severe sanctioning regimen). Given this ambiguity, we established a compliance standard for this item as the percentage of sanctions that are 19 days or less. As is seen in the individual site reports, though, all four DFE sites had mean jail days *per violation* of around a week or less, thus representing substantial congruity with what was found in the Hawaii study. One caveat is that we did not examine any graduation of sanctioning that might have occurred–were subsequent violations sanctioned more harshly than initial violations? Our purpose was simply to examine whether sanctioning overall was "severe."

Item 10 (Sanction Certainty): This item is related to Item # 6, but focuses specifically on the issuance of a sanction for a violation with little or no prospect of escaping the sanction. We assessed the percentage of violations that were issued a specific sanction through the Violation Hearing process.

Item 11 (Sanction Swiftness): This item is related to Item # 7. This item reflects the time in days between the date of the probation Violation Hearing and the date that the sanction (usually confinement) commenced. Sites were expected to allow no more than 3 days to pass between these two dates. We assessed the percentage of sanctions that commenced within the prescribed 3-day window.

Pepperdine University, the training and technical assistance provider to the sites implementing the HOPE model, worked with project coordinators hired by each of the sites. Among other day-to-day project management tasks, each project coordinator was responsible for compiling administrative data about HOPE court probationers from which Pepperdine monitored site-level fidelity to the HOPE model over the course of implementation. Project coordinators accessed official court, corrections, and law enforcement records to abstract select individual-level data and entered these data into an Access database.

These fidelity data include HOPE court activity (warning hearing dates, violation hearing dates), drug testing (date, results, testing frequency), probation violations (date, type), warrant service (date issued, date served, agency), sanctions (date imposed, type, incarceration admission and release date, judges' exceptions), treatment (admission and discharge date, type, setting), and recidivism (arrest date and charge, revocation date and reason, conviction date and offense, incarceration term). These data cover events that occurred from the date of the HOPE warning hearing through the HOPE probation termination date. In addition to their use for fidelity monitoring, these data are also a rich source of information for the outcome and cost evaluations.

To acquire the fidelity data, the evaluation team developed a data transmission protocol that was reviewed and approved by RTI's IRB, and entered a data use agreement with Pepperdine. Pepperdine transmitted two batches of data in April 2014 and November 2014, and a final batch in April 2015. Pepperdine transmitted the fidelity data via an RTI FTP site that encrypts files during uploading. In turn, RTI transmitted the fidelity data to Penn State through RTI's secure FTP portal. As shown in the

individual site reports in Appendix I, data were available for almost all items for all sites, and were generally complete. A few items were not available for a given site; but, sufficient data were available on these items to draw conclusions about fidelity of implementation in all sites.

Depending on the item, analysis was conducted at the client level (Items 2–5) or at the incident level (Items 6–11). For example, Item # 2 (Client Risk Level) was analyzed at the client level. Most probationers assigned to HOPE had at least one risk score although some cases were missing risk data. In cases where multiple scores existed for a given client, we took the score closest to the time of their enrollment into HOPE, giving priority to pre-enrollment scores. Thus, there was one chance for HOPE to achieve fidelity for a given client on that item, by enrolling moderate to high risk offenders in each case (or not). Conversely, Item # 7 (Swiftness of Sanctioning) is measured at the incident level. A HOPE probationer may have had multiple probation violations (incidents) during enrollment in HOPE, thus offering multiple chances for HOPE to achieve fidelity to the model (or not) by bringing that probationer before the HOPE judge for a violation hearing within 3 days of each violation (or not). Thus, on that item, the percentage of violations that were followed by a hearing within the prescribed 3-day window was calculated.

The percentage of compliance was calculated for each of the 11 fidelity items. An overall summative compliance score (i.e., X out of 11) was not calculated. HOPE is a relatively new innovation, and little has been done in the way of process evaluation. Thus, there was a weak basis from which to draw firm conclusions about what sort of summative score would equate to a positive treatment effect. Instead, the reader can examine program performance on each item for each site and draw conclusions about overall fidelity to the model.

Implementation science is still a relatively new field, and much remains to be learned about how well implemented a program needs to be to maximize the likelihood of effectiveness (Proctor, Landsverk, Aarons, Chambers, Glisson and Mittman, 2009). Although there is no consensus on what constitutes “high fidelity” for a given program, there is some empirical basis for conclusions about the degree of match between program design elements and program elements that is associated with good program outcomes. In their meta-analysis of 59 implementation evaluations, Durlak and DuPre (2008) find that fidelity levels of at least 60% can produce measurable, positive program effects.

Esbensen et al. (2011) in their process evaluation of the Gang Resistance Education and Training (GREAT) program employed an implementation checklist like the one used here that measured actual program features against those that were intended. They used a 70% threshold for determining adequate implementation fidelity. They did not, however, offer any empirical or theoretical justification for this threshold; they simply stated that this is the standard they employed.

Finally, the Correctional Program Checklist (CPC) is a widely-used tool that benchmarks a given correctional program against empirically validated principles of effective offender intervention (Lowenkamp, Latessa, and Smith, 2006). The CPC uses a threshold of 65% to qualify a program as being “highly effective” and indicative of strong correspondence between program operations and theories of offender rehabilitation.

Thus, there seems to be some broad support for using a threshold of at least 60% when drawing conclusion about the fidelity of program implementation. ***The discussion of results for the individual sites examines fidelity at the 60% level, and at a higher threshold of 80%***, which the Durlak and DuPre (2008) meta-analysis suggests characterizes a very well implemented program. Thus, a minimum score

of 60% or above on the eleven individual fidelity items was assumed to be indicative that the local HOPE program was implemented with sufficient fidelity as to promote good program performance. To our knowledge, the results presented here represent the first attempt to assess the fidelity of implementation of a HOPE or HOPE-like program in such a rigorous and continuous measurement context.

Documenting the Implementation Experience at the DFE Sites

The rigorous assessment of implementation fidelity, using a checklist and scoring procedure, is a critical component to process evaluation, but is not sufficient for a complete understanding of implementation. Process evaluation must also investigate the rich context of the implementation experience in the practice setting, examining ecological and contextual factors that condition the evolution and achievement (or not) of fidelity (Durlak and DuPre, 2008; Fixsen et al., 2005; Stith, Pruitt, Dees, Fronce, Green, Som and Linkh, 2006). These factors can include barriers and challenges to implementation as well as factors that facilitated implementation; leadership and championship of innovative or novel interventions; group dynamics and communication patterns among intervention stakeholders; stakeholder attitudes towards the intervention; organizational, political and community support and capacity for the intervention; available resources including training and technical assistance; and plans for continuation and sustainability.

We explored and documented these implementation issues through three site visits to each of the four HOPE DFE sites. These visits were timed to correspond to the early period of HOPE implementation (baseline visit), an intermediate period when HOPE was fully operational (interim visit), and final period (final visit) when the site was at or near its client enrollment goals for the DFE. The site research coordinator and/or project coordinator arranged interviews at each round with members of the local HOPE team. The composition of each site's HOPE team differed somewhat due to local administrative variations, but in general the core team consisted of the HOPE judge, HOPE project coordinator, HOPE probation officers, probation management, county jail administrator, and local law enforcement (typically the sheriff). We had good cooperation from most team members with these interviews, thus achieving good saturation. Interviews were conducted by HOPE DFE investigators Gary Zajac, Ph.D. of Penn State University and Debbie Dawes, M.A., of RTI. Zajac and Dawes conducted some interviews together, others separately, depending on local scheduling. Interviews typically lasted 30 to 90 minutes. Each site visit typically lasted 3 to 4 days.

Exhibit 2-10 provides an overview of who was interviewed at each site during each of the three site visits. (BL indicates baseline or first visit, IN indicates interim visit, and FN indicates final visit.) The month and year of each visit is indicated. Bolded entries indicate there was interviewee turnover from one visit to the next. For example, both HOPE probation officers in Arkansas were replaced between the interim and final site visits. The numbers indicate how many individuals with a given role were interviewed. The reasons for changes in these numbers are varied, but usually reflect availability of individuals who played a role in that position. For example, in Arkansas during the baseline visit, the primary HOPE judge was joined in the interview by the backup HOPE judge and another court staffer. During the interim and final visits, the primary HOPE judge was not joined by these individuals. In Texas, probation management was represented by several individuals who assumed prominence during HOPE implementation; thus, the specific individuals interviewed for that role there changed. Zeroes typically indicate that the person was not available during the visit. In Texas, though, the district attorney and

public defender played no role in HOPE and were not interviewed. Similarly, the district attorney in Massachusetts played no role in HOPE. Roles designated as "Other" in Texas included the supervisors of the drug testing unit and the offender assessment unit, and, in Massachusetts, the chief justice of the state superior court, who played an important role in initiating HOPE there. The dynamics of the composition of the HOPE Team are discussed further in the results sections.

Exhibit 2-10. Summary of process evaluation interviews by site and site visit

HOPE Team Member	AR Site Visit Interviews			MA Site Visit Interviews			OR Site Visit Interviews			TX Site Visit Interviews		
	BL 12/12	IN 8/13	FN 10/14	BL 12/12	IN 9/13	FN 8/14	BL 2/13	IN 11/13	FN 7/14	BL 10/12	IN 6/13	FN 10/14
District Attorney	1	2	1	0	0	0	2	2	2	0	0	0
Jail Administrator	0	2	2	1	1	1	1	1	0	0	0	0
HOPE Judge	3	1	1	2	2	2	1	1	1	1	1	1
Probation Management	1	0	0	3	3	3	3	3	3	3	**3**	**3**
HOPE Probation Officers	2	2	**2**	7	7	**7**	2	**3**	3	2	**2**	2
HOPE Program Coordinator	1	1	1	1	0	1	1	**1**	1	1	1	1
Public Defender	1	1	1	1	1	1	1	0	0	0	0	0
Sheriff/Other Law Enforcement	0	0	0	0	1	1	2	2	1	0	1	1
Other	0	0	0	1	1	1	0	0	0	2	2	**2**

Note: BL = baseline; IN = interim; FN = final; **bolded** entries indicate turnover in personnel between visits

The instruments used in the site visits are provided in Appendix A. The completed interviews were analyzed by Dr. Zajac and his research assistants, coding for key implementation themes. During the site visits, Zajac and Dawes also observed violation hearings and warning hearings and met with site data administrators to discuss access to the fidelity data, as well as administrative data needed for the outcome evaluation component of the DFE.

During the final site visits, interviews were also conducted with a small number of HOPE probationers, using the instrument provided in Appendix B. The purpose of these interviews was to obtain their thoughts about HOPE program operations and their experiences participating in the program. The questions were developed from client interview protocols utilized by Zajac in previous correctional program evaluations. Probationers were selected by the research coordinators, and were primarily probationers who had been on HOPE probation for at least 6 months. We interviewed seven HOPE probationers in Arkansas, four in Massachusetts, five in Oregon and five in Texas, for a total of 21 interviews. Personal identifiers were not collected during these interviews.

Finally, also during the final site visit, we conducted a brief examination of the primary drug treatment program used by each site. Each of the DFE sites used drug treatment to some extent. Thus, it

was important to get a least some impression of this treatment to examine potential interactions of treatment with the primary outcomes for HOPE. Each site used multiple drug treatment providers and we could not examine each provider. Instead, we examined the provider at each site that served the largest number of HOPE clients. Zajac conducted this examination using an abbreviated version of the Correctional Program Checklist (CPC; Lowenkamp, et al., 2006), which benchmarks a given correctional program against the principles of effective intervention. Given the short amount of time we were allotted by these providers (typically a 2-hour visit), we were not able to conduct a full CPC evaluation, instead deriving some broad conclusions about the extent to which that program adhered to the basic principles of effective intervention.

Network Data and Analyses

Primary HOPE stakeholders were asked to report on ***communication and involvement*** with each other within the context of the HOPE program during the three visits to each site (fall 2012, summer and fall 2013, and summer and fall 2014). Because all stakeholders reported on all other stakeholders, the resulting data form a (social) network of information. The analytic methods used for these data were social network analysis (SNA; see Butts, 2015, and references therein). Applying these methods presented two challenges: First, the HOPE stakeholder networks were small (analogous to having a small sample size in more traditional analyses), which limited the types and sizes of SNA models that could be utilized. Second, the HOPE stakeholder networks were nearly saturated (analogous to small variances with traditional statistical models), which meant that most stakeholders were connected to most other stakeholders.

Two types of network data were used, those rating communications and those rating involvement/importance:

1. Stakeholder ratings of their level of ***communication*** with each other about the HOPE program, are on a 0-to-4 scale with anchors of "never" and "every day."[12] In the parlance of social network analysis, these ratings "tie" pairs of stakeholders together. This use of the term "tie" should not be confused with tied scores. The rating of stakeholder A's perception of communication with stakeholder B need not match stakeholder B's report of communication with stakeholder A. In fact, stakeholder A can report no communication with stakeholder B (a rating of 0) while stakeholder B could report daily communication with stakeholder A (a rating of 4).
2. Stakeholders perceptions of how involved each stakeholder was in the development of the HOPE program (initial visit) and their importance in the ongoing implementation and operation of HOPE (interim and final visits) are also on a 0-to-4 (not at all involved/important to very involved/important) scale.[13] For analyses using these stakeholder ***involvement/ importance*** data, there was insufficient data for the prosecution and defense stakeholders to be included in the analyses.

Network data can be structured using one of three formats (***Exhibit 2-11)***:

[12] Original responses were scored on a 1 to 5 scale with 1 equal to everyday and 5 equal to never; scores were reverse coded and set to a 0 to 4 scale for analyses.

[13] As with the communications ratings, the original responses were on a 1 to 5 scale and were reverse coded and set to a 0 to 4 scale for analyses.

1. ***Sociomatrix format:*** the rating of each stakeholder (in the rows) is given for each other stakeholder (in columns). Stakeholders do not rate themselves for the communication data, so the diagonal is empty. Although stakeholders did rate their own involvement/ importance, models for self-ratings are not considered here.
2. ***Network graph:*** Ties between stakeholders are indicated by arrows with the absence of an arrow indicating a lack of tie. A single-headed arrow indicates, e.g., one-way communication (as in the above-described example where one stakeholder in a stakeholder pair reports no communication but the other does), and a double-headed arrow indicates that both stakeholders reported at least some communication with each other (a condition called reciprocity). Line widths are proportional to the 0-to-4 rating or the average of ratings when there is reciprocity.
3. ***Edgelist:*** Each pair of stakeholders appears in up to two rows, once for each direction of the tie between stakeholders ('from' and 'to' columns) with the ratings in a separate column.

Some SNA concepts are more easily demonstrated with a sociomatrix and others are better described with an edgelist,[14] while most results are best presented as network graphs.

Four types of SNA methods were used: network graphs, network statistics, stakeholder statistics, and exponential random graph models (ERGM). Each of these is described below.

Network graphs are illustrated in the middle of ***Exhibit 2-11***. These representations of the network include circles called ***nodes or vertices*** that usually represent individual people. The connecting lines are called ***ties, links, or edges*** (hence the name "edgelist" in the bottom panel). Graphical SNA methods can be used with both binary and weighted networks. Binary networks indicate whether two nodes are tied (e.g., whether two stakeholders communicated about HOPE), while weighted networks use data weighted by a rating of the relationship (e.g., how frequently two stakeholders communicate). The thickness of the edges can be proportional to the weights. We use the communication and involvement/importance ratings as weights in the network graphs. In addition, the size of the nodes can be proportional to the stakeholder (or node) statistics, based on how central each is to the network. The distance between nodes, the placing of nodes, and the length of the arrows have no meaning and are based on criteria that visually simplify the overall graph.

Network statistics summarize characteristics of a network. ***Density*** is a basic descriptive measure of the saturation of a network with ties and is computed by dividing the number of reported ties by the total possible number of ties. ***Clustering*** is based on connections among triplets, rather than pairs as is done with the density measure. A closed triplet has three stakeholders who are connected by single or double headed arrows. The network graph in ***Exhibit 2-11*** is an example of a closed triplet. An open triplet includes at least one pair of stakeholders with no tie between them (e.g., if one of the edges were removed in ***Exhibit 2-11***). There are two general types of ***clustering measures***. The first is a binary (or unweighted) measure that accounts for the presence or absence of links between stakeholders. The second is a weighted measure that incorporates ratings to account for the strength of connections

[14] Some SNA software packages require data as sociomatrices while other software requires edgelists. For large networks, edgelists are more computationally efficient.

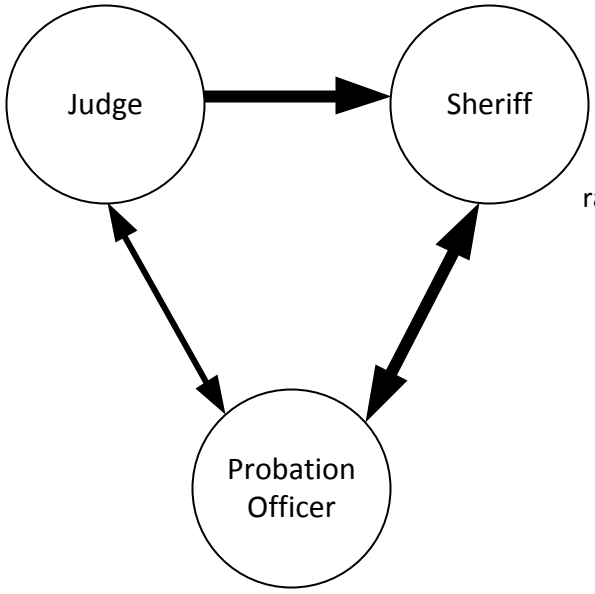
between people.[15] If the strength of connections (or weights) are randomly distributed in a network, the weighted measures will be equal to the binary measure. When a weighted measure is greater than the binary measure, it indicates that higher communication or involvement/importance ratings tend to appear among people in closed triplets than among those in open triplets.

Exhibit 2-11. Hypothetical social network data in three formats

Sociomatrix format

	Judge	Sheriff	Probation Officer
Judge	-	4	4
Sheriff	0	-	4
Probation Officer	1	4	-

Network graph format



Note: The single headed arrows indicate that the judge reported communicating with the sheriff about HOPE, but the sheriff did not report communicating with the judge about HOPE. Arrow thickness is proportional to the rating, or average rating in the case of two headed arrows.

Edgelist format

From	To	Rating
Judge	Sheriff	4
Judge	Probation Officer	4
Sheriff	Judge	0
Sheriff	Probation Officer	4
Probation Officer	Judge	1
Probation Officer	Sheriff	4

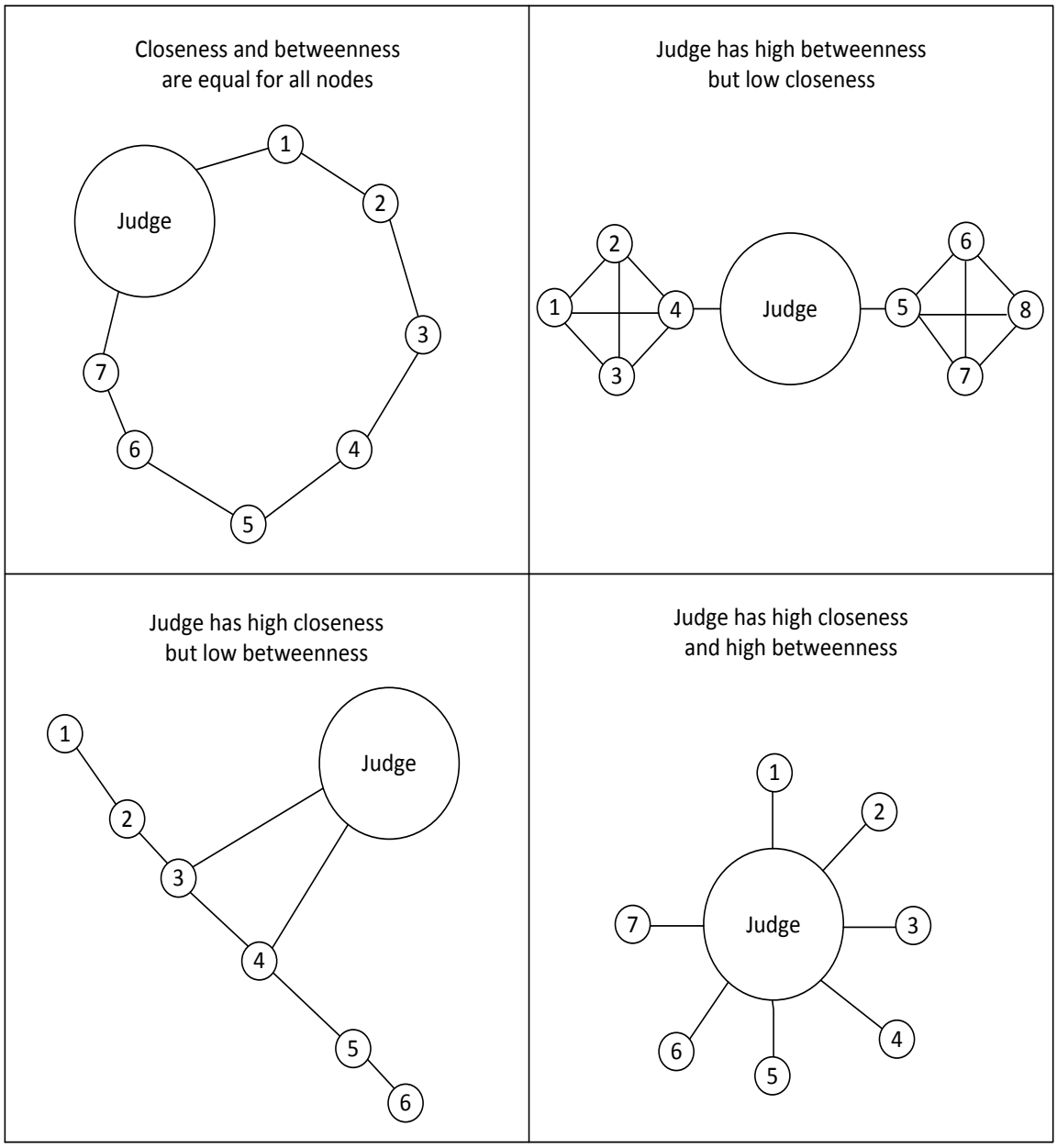
[15] The small networks reported herein are close enough to saturated (i.e., most stakeholders communicate with most other stakeholders) that many SNA methods would not be feasible in the absence of communication ratings (i.e., if we simply recorded whether stakeholders communicated rather than the frequency of communication). Researchers who want to apply SNA methods to evaluations of small networks should keep this limitation in mind and design data collection instrumentation to ensure that weighted network data are obtained.

The most commonly used ***node statistics*** are centrality measures, which, for our data, are computed for each stakeholder at each interview and describe how central each stakeholder is to communication in the network. We use weighted versions of centrality measures where the communication or involvement/importance ratings are considered when computing centrality measures. We used two measures of centrality: ***betweenness centrality*** and ***closeness centrality***. ***Exhibit 2-12*** illustrates four hypothetical networks. In the lower left panel, the shortest path (the number of ties to get from one node to another) from node 1 to node 6 is along the 5 ties between nodes 1 through 6. In the upper left panel, the shortest path from node 1 to node 6 is via the judge node and node 7. ***Betweenness centrality*** is a measure of how often a stakeholder appears on the shortest paths between the other stakeholders in the network. In the four panels of ***Exhibit 2-12***, the judge has varying levels of betweenness centrality. In the upper right panel, the judge is a bottleneck between two sub-networks (nodes 1 through 4 and nodes 5 through 8) and has high betweenness centrality but has lower closeness centrality than nodes 4 and 5 that are closely connected to many other nodes. ***If the judge-driven HOPE model holds in practice, the betweenness centrality should be large for the judge***, but large betweenness centrality may also be observed for other stakeholders.

Closeness centrality is a measure of the inverse of how far a node is from all other nodes via shortest paths. A stakeholder with a high closeness centrality score is connected to the rest of the network via shorter paths than is a stakeholder with a low closeness centrality score. In the lower left panel of ***Exhibit 2-12***, the judge is close to the remaining nodes in that their position is centered and hence relatively closer to the rest of the stakeholders. All nodes except 1 and 6 have two ties, but the judge and nodes 3 and 4 have higher closeness centrality than nodes 2 or 5 because they are centered in the network. In contrast, the judge has low betweenness centrality since their position is not between other stakeholders. ***If the judge-driven HOPE model holds in practice, the closeness centrality should be large for the judge as the judge theoretically drives connections in the network.*** This is idealized in the lower right panel of ***Exhibit 2-12***.

In summary, ***betweenness centrality measures how intermediate a stakeholder is to connections between other stakeholders, while closeness centrality measures how closely connected a stakeholder is to many other stakeholders.*** With the small networks considered here, most stakeholders are reciprocally connected to most other stakeholders. In the context of the communication ratings, a large betweenness centrality measure indicates those who mediate the most communication exchanges about HOPE between other stakeholders, while a large closeness centrality measure indicates those who tend to communicate with the most stakeholders about HOPE. In the context of involvement/ importance ratings, a large betweenness centrality measure indicates those many rate as very involved or important, while a large closeness centrality measure indicates those who tend to be highly involved with the other stakeholders. We incorporated these stakeholder statistics into the network graphs by making the size of the circle proportional to the size of each stakeholder's betweenness centrality and closeness centrality.

Exhibit 2-12. Comparisons of betweenness and closeness centrality using hypothetical HOPE networks

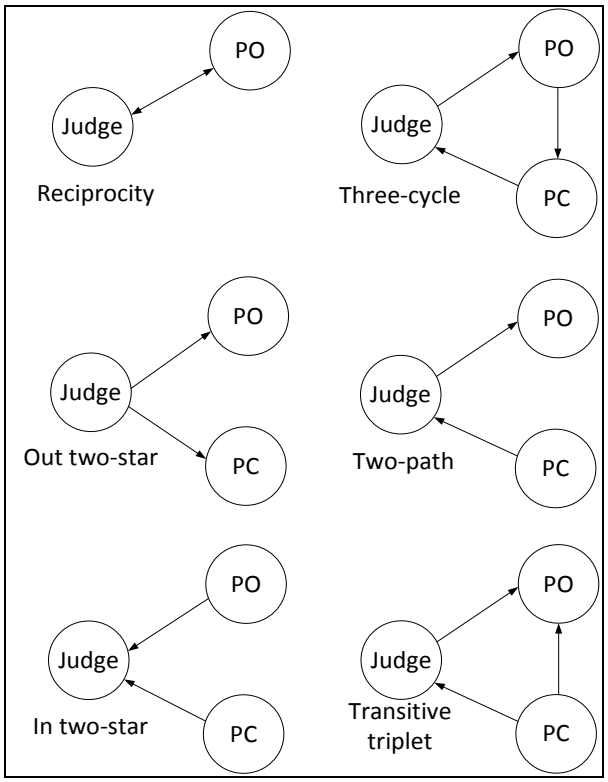


The ***Exponential Random Graph Model (ERGM)*** is an extension of logistic regression, where the dependent variable is the presence or absence of a tie between two stakeholders and the data are structured as an edgelist (bottom panel of ***Exhibit 2-11***). Extensions of the model allow for weighted data (e.g., the communication and involvement ratings). The independent variables of an ERGM are characteristics of the network.[16] Network characteristics used as predictors are illustrated in the left

[16] Predictors could also be characteristic of the stakeholder. Through the interview process, we collected data on each stakeholder's opinion of the HOPE program and coded it as either positive or negative. The results were almost unanimously positive, which led to predictors with (near) zero variance, and this precluded their use in the ERGMs.

panel of ***Exhibit 2-13*** as hypothetical social network sub-diagrams of two or three stakeholders.[17] The first network characteristic is "reciprocity," which represents perceived two-way communication and indicates whether the stakeholders in a pair both reported at least some level of communication. In the hypothetical example of ***Exhibit 2-11***, the judge reported communicating with the sheriff, but the sheriff did not report communicating with the judge, so the communication is not reciprocal (these pairs would receive a score of 0 for reciprocity). An "*in* two-star" network characteristic is present when one stakeholder reports *receiving* information from two other stakeholders who do not report communicating with each other. Alternately, when a stakeholder reports *sending* information to two other stakeholders who do not report communicating with each other, this is referred to as an "*out* two-star." A "two-path" network characteristic is present when information flows through a central stakeholder. A transitive triplet is present when a stakeholder reports receiving both direct communications from another stakeholder and indirect communication from that stakeholder via a third stakeholder (e.g., the probation officer receives information directly from the project coordinator and indirectly through the Judge). A three-cycle measure indicates the extent to which communication "cycles" through three stakeholders, as is seen in the Judge -> Sheriff -> Probation Officer -> Judge three-cycle representation in the middle panel of ***Exhibit 2-11***.

Exhibit 2-13. Hypothetical exponential random graph models of HOPE stakeholders



NOTE: PO = probation officers. PC = project coordinator. PM = probation manager.

[17] In this example, when a characteristic is one of the columns in the illustrative edgelist format of Exhibit 2-11, we note this. However, the absence of the characteristic as one of the columns in the edgelist does not imply that the hypothetical data are devoid of such a characteristic, we simply limited the number of illustrative characteristics for simplicity.

To clarify these concepts, the edgelist from ***Exhibit 2-11*** is replicated in ***Exhibit 2-14*** with columns added to show reciprocity and three-cycle measures to provide a sense of how the data are structured for the ERGM. For simplicity, the other network characteristics illustrated in the left panel of ***Exhibit 2-13*** are not shown. As noted above, because the data in the rows of the edgelist are not independent, one cannot simply use logistic (or multinomial logistic) regression to predict ties (or ratings) from the characteristics just discussed. To deal with this lack of independence, ERGMs are fit using Markov chain Monte Carlo methods (Snijders, 2002).

Exhibit 2-14. Hypothetical edgelist with ratings, reciprocity, and three-cycle measures

From	To	Rating	Reciprocity	Three cycle
Judge	Sheriff	4	0	1
Judge	Probation Officer	4	1	0
Sheriff	Judge	0	0	0
Sheriff	Probation Officer	4	1	1
Probation Officer	Judge	2	1	1
Probation Officer	Sheriff	4	1	0

To more concretely show how the structures in the left panel of ***Exhibit 2-13*** exist in the current data, we examine the project coordinator (PC), probation manager (PM), and judge relationships from the initial interview for the Arkansas site (right panel of ***Exhibit 2-13)***. For simplicity, the lines simply denote relationships and are not proportional to the ratings data, and the circles are all fixed to the same size (i.e., are not proportional to a centrality measure). Most of the structures in the left panel of ***Exhibit 2-13*** are present simultaneously in actual data in the right panel. All structures present in the left panel except the three-cycle are present in the actual data in the upper left corner of the right panel. It is challenging to visually pick out the individual structures of the left panel as illustrated in the right panel 3. Hence, using the Bayesian Information Criterion (BIC) criterion to select the best fitting structure facilitates the understanding of which structure best describes the overall data.

The predictors in an ERGM model are known as ***difference statistics*** (defined in Lubbers & Snijders, 2007) and are calculated for each stakeholder pair. Difference statistics are based on the count of network characteristics (e.g., the number of reciprocal relationships, the number of three-cycles, etc.) when the tie between a pair of stakeholders is forced to be present minus the count of network characteristics when the tie between a pair of stakeholders is forced to be absent. One can imagine adding columns of difference statistics to the left of the hypothetical data in ***Exhibit 2-14*** (see also the Methodological Appendix of Laven, Krymkowski, Ventriss, Manning, & Mitchell, 2010).

When fitting an ERGM to data, each difference statistic has an estimated coefficient and an associated standard error. The null hypothesis is that the network characteristics of a given type (e.g., reciprocity, three-cycles) occur no more or less than expected by random chance (Morris, Handcock, & Hunter, 2008) in a network of the same density (Laven, Krymkowski, Ventriss, Manning, & Mitchell, 2010). A reciprocity characteristic with a hypothetical coefficient of 0.405 as an example would be interpreted like a logistic regression coefficient: for a given pair of stakeholders, a one-unit increase in

the difference statistic for reciprocity leads to a change of 0.405 in the log-odds of a tie leading to reciprocity being present. Similar to computing the odds ratio for logistic regression, we compute exp(0.405)=1.5, indicating we are 1.5 times as likely to see a tie leading to reciprocity than expected by chance. When the coefficient is positive and significant, we are more likely to see ties that would lead to reciprocity when the reciprocity difference statistic is 1 than when it is 0. If the coefficient is negative and significant, we are less likely to see ties that would lead to reciprocity when the reciprocity difference statistic is 1 than when it is 0.

We fit the ERGMs using the R (R Core Team, 2015) package statnet (Handcock, Hunter, Butts, Goddreau, & Morris, 2003) using one independent variable (one of the network characteristics illustrated in the left panel of ***Exhibit 2-13***) at a time. Due to the small network size, models including multiple independent variable would not converge. For weighted networks, we use the approach of Krivitsky (2012). Rather than report on each of the six characteristics in the left panel of ***Exhibit 2-13*** for each network at each interview, we report which characteristic fit the network best. We selected the best fitting network characteristic by comparing the BIC across the six models for each network.

One challenge of current SNA analytic methods is that they are unable to distinguish between missing data and lack of a tie. Lack of tie is indicated by a zero (0) communication rating (or involvement/importance rating), in our data indicating "never" for the communication rating and "not at all involved/important" for the involvement/importance rating. In practice, the third row of the edgelists of ***Exhibit 2-11*** and ***2-14*** would be excluded so that the SNA software would correctly identify this as a lack of communication between stakeholders. However, this is also currently the only option for designating missing data, and hence the two are confounded. Therefore, we addressed the issue of missing data using within-person imputation by either carrying the last observation forward (if an observation was missing at later interviews) or using the mean of the other observations (if the first observation was missing). When data were missing for all three time points, those who collected the data indicated it was their impression from qualitative data that it was reasonable to assume no communication was occurring. This occurred only for the prosecutor in Massachusetts and the prosecutor and defense in Texas.

2.3. Outcome Evaluation

The outcome evaluation was designed around the four-site randomized controlled trials to address the following research questions:

1. Does HOPE participation improve compliance with conditions of supervision and reduce violations?
2. Does HOPE participation reduce recidivism, measured by arrest, conviction, and probation revocation?
3. What is the impact of HOPE on jail days served and prison days sentenced?
4. What is the impact of HOPE on drug use?
5. Does HOPE participation change potential mediators including dynamic recidivism risk factors such as employment and housing stability?
6. Does HOPE participation change attitudes that are potential mediators, including participants' criminal thinking/attitudes, perceptions of locus of control, and perceptions of the criminal justice system fairness/legitimacy?

Outcome data were collected from multiple sources, including up to three in-person interviews (baseline and 6- and 12-months post-baseline) with each study participant, telephone "mini-interviews" with a randomly selected subsample of those who consented to baseline interviews, oral swab drug tests of a randomly selected subsample of individuals interviewed post-baseline in the community, and local and state administrative agency data sources.

A variety of analytic techniques were used to address the research questions. These included descriptive statistics to describe the characteristics of our study populations, survival models to examine time to events (e.g., violation, new arrest) and factors associated with time between events (e.g., violations and positive drug tests), and count models to examine the factors associated with numbers of recidivism events.

Data Collection Protocols and Procedures

Full-time research coordinators were hired in the four sites by the evaluation to be available to introduce the study to HOPE-eligible probationers, perform evaluation intake, administer the baseline interview, and conduct follow-up interviews at 6 and 12 months after the baseline interview. These individuals also helped with the collection of administrative data and, as noted earlier, arranged interviews for the implementation and process evaluation. The research coordinators worked in offices provided by the probation department in three sites and in an office rented for the study across from the courthouse in the fourth site.

RESEARCH COORDINATORS

Research coordinators hired by the evaluation conducted interviews, observed warning and violation hearings, and assisted with administrative data collection is the DFE sites.

Interviews (baseline, 6M, and 12M)

Procedures. Interviews were administered using Audio-Computer-Assisted-Self-Interviewing (ACASI) technology through which the respondents listened to each question through headphones and entered responses into a laptop computer. The same instrument (with necessary modifications such as time-period references) was used for all waves of data collection. The instrument is included in Appendix C. Questions in the instruments were derived from those used in other studies with similar populations and included the following domains (References for *items/scales in italics* are in ***Exhibit 2-15.)***:

- Demographics (Age, Race, Acculturation)
- Education (Education Status)
- Military Experience (Military Experience)
- Employment (Sources of Support and Employment Status, Reasons for Not Working, Most Recent Job—Finding the Job and Job Characteristics, Wages and Benefits, Other Financial Support)
- Housing (Type of Housing and Household Composition, Housing Stability, *Neighborhood Information*)
- Family Background (Marital/Partner Status, Children, Family Affiliation, Family Criminal History, *Family Emotional Support*, *Quality of Intimate Partnership*, *Relationship with Children* (questions only asked of R's with children under 18))

- Peer Relationships (*Peer Criminal Behavior, Friendships*)
- Program Operations and Services (Service Need, Services Received—Financial, Healthcare, and Legal Assistance, Services Received—Other Services, Medical (Physical Health) Care, Mental Health Care, AOD Treatment)
- Physical and Mental Health (Physical Health Status (overall), Mental Health Status (overall), Mental Health Treatment, Mental Health Hospitalizations)
- Crime and Delinquency (Criminal History – lifetime arrests and incarcerations, Supervision Status and Officer Contacts, Supervision Conditions and Violations (asked only if R is not incarcerated), Probation Sanctions/Rewards)
- Attitudes (*Self-Efficacy*, *Readiness for Change, Legal Cynicism*, *Substance Abuse Treatment Motivation, Community Involvement*)
- T-ACASI Domains (asked on ACASI instrument and T-ACASI) (*Attitude Toward Supervision Officer*; *Deterrence Related to Drug Use; Deterrence Related to Compliance with Conditions*; *Perceived Fairness of Sanctions*; *CSS-M Attitudes Toward the Law Subscale; CSS-M Tolerance for Law Violations*; *CSS-M Identification with Criminal Others*; Attitude Toward Judge, Deterrence—Likelihood, and Severity of Sanctions for Rule Violations; Locus of Control)

Sources and citations are shown in ***Exhibit 2-15***. Most of the scales were previously used in other multi-site studies of offenders conducted by RTI, the Urban Institute and others, including the Urban Institutes Returning Home project, the Serious and Violent Offender Reentry Initiative (SVORI; Lattimore and Visher, nd) Multisite Evaluation conducted by RTI and Urban, and the Multisite Adult Drug Court Evaluation conducted by Urban, RTI, and the Center for Court Innovation.

Usability testing was conducted with five probationers who were receiving services at the Durham (NC) Criminal Justice Resource Center. These individuals took the survey and provided feedback on the ease of use and understandability of the instrument. These individuals were provided $40 to compensate for their time. The final instrument took approximately 30 minutes to complete.

The evaluation research coordinators were trained in consent procedures, interview administration, and tracing (for follow up) activities. All procedures were reviewed and approved by an RTI Institutional Review Board. Procedures were also reviewed and approved by NIJ's Human Subjects Protection Officer.

Each probationer identified as HOPE-eligible was referred to the evaluation research coordinator as described earlier. The coordinator entered information from the referral slip into a spreadsheet and then described the study and asked the individual if he or she would be willing to complete a baseline interview. The research coordinator reviewed the consent form (see Appendix D for an example) with the individual and, if the individual consented, the research coordinator showed the individual how to use the ACASI interview to complete the computerized baseline interview.[18]

The research coordinators assured that respondents had privacy so that their responses could not be viewed by anyone. Interviews were conducted in the research coordinators' offices (located in probation offices in three sites; after, being unable to find suitable space in the Saline County probation offices or court, in a rented office across from the court house; in local jails; and in suitable places out in the community (e.g., individuals' homes or libraries). Follow-up interviews were arranged by the

[18] If the individual did not want to participate in the interview, the research coordinator thanked the individual and sent the probationer back to the probation office.

research coordinators and were conducted in the same types of locations as the baseline interview. ACASI interview data were transmitted to RTI servers each evening and processed to a SAS data base.

Exhibit 2-15. Outcome evaluation instrument sources for measures and scales

Domain	Study Source	Citation
Neighborhood Information	Returning Home (& SVORI)	Urban Institute, Returning Home (http://www.urban.org/policy-centers/justice-policy-center/projects/returning-home-study-understanding-challenges-prisoner-reentry)
Family Emotional Support	Returning Home (& SVORI)	Urban Institute, Returning Home
Quality of Intimate Partnership	Returning Home (& SVORI)	Urban Institute, Returning Home
Relationship with Children	Returning Home (& SVORI)	Urban Institute, Returning Home
Peer Criminal Behavior	Returning Home (& SVORI)	Urban Institute, Returning Home
Friendships	Returning Home (& SVORI)	Urban Institute, Returning Home
Self-Efficacy	Returning Home (subset used in SVORI); Subscale of the TCU Correctional Residential Treatment Self-rating at intake, Self-Efficacy Scale (items 92, 97, 95, 94)	"Mastery Scale" (Pearlin & Schooler, 1978).
Readiness for Change	Returning Home (& SVORI)	Urban Institute, Returning Home
Legal Cynicism	Returning Home (& SVORI)	Urban Institute, Returning Home Sampson & Bartusch, 1998
Substance Abuse Treatment Motivation	RTI's Drug Court Evaluation Self-rating at intake, Desire for Help scale (items 7, 29, 40, 58, 86, 80 [part of the treatment readiness scale], 8 [part of external pressures scale]) The last question is not part of the TCU.	Subscale of the TCU: http://www.ibr.tcu.edu/pubs/datacoll/cjforms.html
Community Involvement	HOPE VI Panel Study (RTI) modified for SVORI (from a civic action scale)	Civic Attitudes and Skills Questionnaire, (Moely, Mercer, Ilustre, Miron, & McFarland, 2002)
Attitude Toward Supervision Officer	Multisite Adult Drug Court Evaluation	Center for Court Innovation in 2000 (unpublished)
Deterrence Related to Drug Use	Multisite Adult Drug Court Evaluation	Center for Court Innovation in 2000 (unpublished)
Deterrence Related to Compliance with Conditions	Multisite Adult Drug Court Evaluation	Center for Court Innovation in 2000 (unpublished)
Perceived Fairness of Sanctions	SVORI	No earlier reference found
Criminal Sentiments Scale-Modified (CSS-M)		Criminal Sentiments Scale-Modified (Simourd, 1997)

Respondent Compensation. The evaluation team conducted negotiations with each site with respect to compensation payments to individuals who agreed to participate in the interviews. The original proposal was to provide a nominal amount—a $5 gift card or equivalent. It quickly became apparent that the probationers were not willing to complete a 30-minute interview for that amount—and new negotiations were conducted with the sites to increase the amount to $20 or equivalent. The issue of providing any form of compensation to the probationers was a problem with some judges and departments and resulted in a series of negotiations and differences in the types of compensations we could offer probationers who were interviewed in the community. No compensation was offered to individuals who were incarcerated at the time of an interview (mostly during follow-up interviews). For respondents in the community, the compensation was as follows:

- Saline County, Arkansas: $20 gift card
- Essex County, Massachusetts: $20 cash
- Clackamas County, Oregon: $20 cash
- Tarrant County, Texas: $5 gift card or $20 probation fee credit[19]

Oral Swab Drug Tests. In addition, individuals who completed 6- and 12-month follow-up interviews were sometimes asked to provide an oral swab for submittal for drug testing.[20] These tests provided the only means for the study to determine whether HOPE was reducing drug use compared to PAU, as there was extensive random drug testing of HOPE participants and little to no random testing of individuals in PAU. Thus, test results from administrative records were not comparable for the HOPE and PAU groups. Individuals who agreed to provide a swab were provided a $5 gift card.[21]

Interview Response Rates.

As is always the case, not everyone completed interviews. Some individuals refused; others could not be located or contacted for follow-up interviews. For this study, there were other circumstances that resulted in interviews not being conducted. These include the following:

- At baseline, in some cases the research coordinator was not available to conduct an interview at the time the HOPE-eligible individual was identified. In these cases, if the interview could be arranged within a few days, the interview was conducted even though random assignment would have already occurred. In most cases, however, the interview was simply skipped.
- In Arkansas, the evaluation research coordinator position was unfilled for 2 months (mid-June to September 9, 2013) due to staff turnover resulting in some missed baseline interviews.
- For follow-up interviews, research coordinators were sometimes stymied in conducting interviews in the community because they needed to be in their offices in case new HOPE-

[19] Eventually, the $5 gift card was dropped and the only offer in Texas was a $20 probation fee credit; the fee credit had been the choice of more than 90% of respondents who had been offered the choice.

[20] During the initial year or so of the evaluation, a schedule was provided to each site that indicated random weeks that had been designated to offer those completing follow-up interviews the opportunity to provide an oral swab. A 50% random selection process was used. Towards the end of the study, everyone was asked to provide a swab. The drug test analyses were conducted through a contract with Omega Labs.

[21] In Tarrant County, compensation for providing an oral swab was eventually changed to offer a $5 probation fee credit instead of the gift card. Individuals who were no longer on probation were offered the gift card.

eligible cases were referred to them. The original design had assumed that evaluation intake would be completed in roughly 9 months' time—meaning that the baseline interviews would be mostly complete before the research coordinators needed to go out into the community to conduct 6-month follow-up interviews. In fact, as noted earlier, enrollment took considerably longer than 6 months—lasting 16 months in Oregon, 22 months in Massachusetts, and 26 months in Arkansas and Texas. This reduced the ability of the research coordinators to conduct follow-up interviews.

Exhibit 2-16 shows the interview results for the baseline (Wave 1) interview. Among 1,504 probationers in the study sample, 978 (65%) completed a Wave 1 interview. HOPE and PAU probationers were equally likely to complete an interview (66% and 64%, respectively). The remaining cases included 306 (20%) probationers who refused the interview, 91 (6%) who provided consent to be interviewed later but did not return to the research coordinator's office to complete the interview, 90 (6%) where the research coordinator was not available to administer the interview, and 39 (3%) who were not interviewed for other reasons.

Exhibit 2-16. Wave 1 (baseline) interview status by group

	HOPE		PAU		ALL CASES	
	N	%	N	%	N	%
Interview complete	491	66.1	487	64.0	978	65.0
Interview refused	145	19.5	161	21.2	306	20.4
Consented, no return for interview	49	6.6	42	5.5	91	6.0
Research coordinator not available	41	5.5	49	6.4	90	6.0
Survey ineligible (age, language restriction)	8	1.1	8	1.1	16	1.1
Unable to contact (homeless, moved away)	4	0.5	7	0.9	11	0.7
Incarcerated/access denied	2	0.3	6	0.8	8	0.5
Other	3	0.4	1	0.1	4	0.3
TOTAL STUDY ELIGIBLE CASES	743	100	761	100	1,504	100

There were no differences in interview status between those assigned to HOPE and those assigned to PAU.

Wave 1 interview completion varied across the sites (***Exhibit 2-17***). Interview completion was highest in Massachusetts (88%) and lowest in Arkansas (50%). Massachusetts had the lowest refusal rate (9%), and Arkansas and Oregon had the highest rates of refusal (25%). About 15% of interviews in Arkansas and 10% of interviews in Texas were not completed because the research coordinator was not available (both sites experienced turnover in this position). Survey ineligibility due to age (under 18 years old) or language restrictions was a small issue in Texas, where 4% of cases were not interviewed for these reasons. Within study sites, ***there were no significant differences with respect to Wave 1 interview status between HOPE and PAU probationers in Massachusetts or Texas. In Arkansas, HOPE probationers were more likely than PAU probationers to complete a Wave 1 interview*** (57% versus 42%), and PAU probationers were more likely than HOPE probationers to be refusers (32% versus 19%).

Exhibit 2-17. Wave 1 (baseline) interview status by site

Interview Status	AR		MA		OR		TX	
	N	%	N	%	N	%	N	%
Interview complete* **	170	49.7	346	88.3	245	62.2	217	57.7
Interview refused* **	87	25.4	37	9.4	99	25.1	83	22.1
Consent, no return for interview* **	23	6.7	6	1.5	40	10.2	22	5.9
Research coordinator not available* **	50	14.6	1	0.3	2	0.5	37	9.8
Survey ineligible (age, language restriction)* **	1	0.3	0	0	0	0	15	4.0
Unable to contact (homeless, moved away) **	7	2.1	2	0.5	2	0.5	0	0
Incarcerated/access denied	3	1	0	0	4	1	1	0.3
Other	1	0.3	0	0	2	0.5	1	0.3
TOTAL STUDY ELIGIBLE CASES	342	100	392	100	394	100	376	100

***Interview status differed across sites ($p < 0.001$); **interview status differed across sites ($p < 0.01$)

Attempts were made to recruit all 1,504 probationers for a Wave 2 (6-month follow-up) interview. Interviews were completed by 536 (36%) probationers (***Exhibit 2-18***). ***HOPE probationers were more likely than PAU probationers to complete a Wave2 interview (41% versus 30%).*** Among those who were not interviewed, HOPE probationers were more likely than PAU probationers to be incarcerated (11% versus 8%). PAU probationers were more likely than HOPE probationers to be unable to contact (32% versus 24%), and to have an active warrant (10% versus 6%). The refusal rate was low (5%); HOPE and PAU probationers were equally likely to refuse a Wave 2 interview.

Exhibit 2-18. Wave 2 (6-month) interview status by group

	HOPE		PAU		ALL CASES	
	N	%	N	%	N	%
Interview complete* **	304	40.9	232	30.5	536	35.6
Unable to contact (homeless, moved) **	176	23.7	243	31.9	419	27.9
Research coordinator not available	85	11.4	105	13.8	190	12.6
Incarcerated/access denied*	80	10.8	58	7.6	138	9.2
Active warrant**	46	6.2	75	9.9	121	8.1
Interview refused	35	4.7	36	4.7	71	4.7
Other	6	0.8	8	1.1	14	0.9
Consent, no return for interview	6	0.8	2	0.3	8	0.5
Survey ineligible (age, language restriction)	4	0.5	1	0.1	5	0.3
Deceased	1	0.1	1	0.1	2	0.1
TOTAL STUDY ELIGIBLE CASES	743	99.9	761	100.0	1,504	99.9

*** HOPE and PAU differ at $p < 0.001$; ** HOPE and PAU differ at $p < 0.01$; * HOPE and PAU differ at $p < 0.05$.

Wave 2 interview status varied across the study sites (***Exhibit 2-19***). Wave 2 interview completion was highest in Massachusetts (44%) and lowest in Texas (24%). Inability to contact probationers for an

interview was most problematic in Massachusetts and Oregon (40% and 32%). About 20% of probationers in Arkansas and Texas did not complete an interview because the research coordinator was not available. Incarceration as the reason for a noninterview was highest in Texas (20%) and lowest in Massachusetts (3%). Having an active warrant as the reason for a noninterview was highest in Oregon (16%) and lowest in Massachusetts (2%). ***Within sites, there were no significant differences with respect to Wave 2 interview status between HOPE and PAU probationers in Massachusetts or Oregon. In Arkansas, HOPE probationers were more likely than PAU probationers to complete an interview*** (52% versus 31%). In Texas, HOPE probationers were more likely than PAU probationers to be incarcerated (25% and 15%), and PAU probationers were more likely than HOPE probationers to have an active warrant (14% versus 6%).

Exhibit 2-19. Wave 2 (6-month) interview status by site

Interview Status	AR		MA		OR		TX	
	N	%	N	%	N	%	N	%
Interview complete* **	145	42.4	172	43.9	128	32.5	91	24.2
Unable to contact (homeless, moved away)* **	77	22.5	155	39.5	127	32.2	60	16.0
Research coordinator not available* **	69	20.2	31	7.9	11	2.8	79	21.0
Incarcerated/access denied* **	13	3.8	10	2.6	39	9.9	76	20.2
Active warrant* **	14	4.1	6	1.5	63	16.0	38	10.1
Interview refused	13	3.8	16	4.1	17	4.3	25	6.7
Other	3	0.9	1	0.3	7	1.8	3	0.8
Consent, no return for interview* **	7	2.1	0	0.0	1	0.3	0	0.0
Survey ineligible (age, language restriction)*	1	0.3	0	0.0	0	0.0	4	1.1
Deceased	0	0.0	1	0.3	1	0.3	0	0.0
TOTAL STUDY ELIGIBLE CASES	342	100	392	100	394	100	376	100

***Interview status differed across sites ($p < 0.001$); **interview status differed across sites ($p < 0.01$); *interview status differed across sites ($p > 0.05$).

Attempts were made to recruit the 1,312 probationers who were eligible to complete a Wave 3 (12-month follow-up) interview.[22] Interviews were completed by 459 (35%) probationers (***Exhibit 2-20***). ***HOPE probationers were more likely than PAU probationers to complete an interview*** (41% versus 29%). Similar to results for Wave 2, PAU probationers were more likely than HOPE probationers to be unable to contact (35% versus 27%), and to have an active warrant (11% versus 6%).

[22] Some individuals had not been in the study 12 months when the evaluation ended interview data collection.

Exhibit 2-20. Wave 3 (12-month) interview status by group

	HOPE		PAU		ALL CASES	
	N	%	N	%	N	%
Interview complete* **	267	41.0	192	29.1	459	35.0
Unable to contact (homeless, moved away) **	176	27.0	229	34.6	405	30.9
Incarcerated/access denied	74	11.4	62	9.4	136	10.4
Research coordinator not available	52	8.0	67	10.1	119	9.1
Active warrant **	39	6.0	71	10.7	110	8.4
Interview refused	30	4.6	29	4.4	59	4.5
Other	7	1.1	8	1.2	15	1.1
Deceased	4	0.6	3	0.5	7	0.5
Consent, no return for interview	2	0.3	0	0.0	2	0.2
TOTAL STUDY ELIGIBLE CASES[1]	651	100.0	661	100.0	1,312	100.1

[1]Excludes 192 (12.8%) study subjects who had not been enrolled in the study for at least 12 months and, therefore, were not eligible for a 12-month interview when the study ended.
*** HOPE and PAU differ at $p < 0.001$; ** HOPE and PAU differ at $p < 0.01$.

Across study sites, Wave 3 interview status varied (***Exhibit 2-21***). The rate of interview completion was highest in Arkansas (54%) and lowest in Texas (26%). Being unable to contact probationers for an interview ranged from a high of 47% in Massachusetts to a low of 16% in Texas. Incarceration as the reason for a noninterview was highest in Texas (20%) and lowest in Massachusetts (5%). Probationers not completing an interview because the research coordinator was not available ranged from 21% in Texas to 1% in Oregon. Having an active warrant as the reason for no interview was highest in Oregon (17%) and lowest in Massachusetts (3%).

Exhibit 2-21. Wave 3 (12-month) interview status by site

Interview Status	AR		MA		OR		TX	
	N	%	N	%	N	%	N	%
Interview complete* **	137	53.5	118	35.9	116	29.4	88	26.4
Unable to contact (homeless, moved away)* **	47	18.4	156	47.4	147	37.3	55	16.5
Incarcerated/access denied* **	15	5.9	15	4.6	38	9.6	68	20.4
Research coordinator not available* **	29	11.3	16	4.9	3.0	0.8	71	21.3
Active warrant* **	12	4.7	10	3.0	68	17.3	20	6.0
Interview refused **	11	4.3	10	3.0	9	2.3	29	8.7
Other **	2	0.8	0	0.0	11	2.8	2	0.6
Deceased	1	0.4	4	1.2	2	0.5	0	0.0
Consent, no return for interview*	2	0.8	0	0.0	0	0.0	0	0.0
TOTAL STUDY ELIGIBLE CASES[1]	256	100	329	100	394	100	333	100

[1] Excludes 192 (12.8%) study subjects who were not eligible for a 12-month interview when the study ended.
*** HOPE and PAU differ at $p < 0.001$; ** HOPE and PAU differ at $p < 0.01$; * HOPE and PAU differ at $p < 0.05$.

Within sites, ***there were no significant differences with respect to Wave 3 interview status between HOPE and PAU in Massachusetts. In Texas and Oregon, HOPE probationers were more likely than PAU probationers to complete an interview*** (38% and 34% versus 15% and 25%). Among those who were not interviewed, HOPE probationers in Arkansas were more likely than PAU probationers to be incarcerated (9% versus 3%). PAU probationers in Arkansas and Texas were more likely than their HOPE counterparts to be unable to contact (25% and 23% versus 13% and 10%). PAU probationers in Oregon and Arkansas were more likely than HOPE probationers in these sites to have an active warrant (22% and 8% versus 13% and 1%). In Texas, more PAU probationers than HOPE probationers did not complete an interview because the research coordinator was not available (26% versus 17%).

Interview Data Response Bias

We tested the data to examine for three types of response bias: (1) Were there differences between those who consented for baseline interviews and those who did not? (2) Were there differences between HOPE and PAU on the baseline interviews? And (3) Were there differences among those who completed a follow-up interview and those who did not?

The following variables from intake or administrative sources were tested to see whether they predicted whether a participant consented to participate in the interview component of the study (baseline): gender, age at intake, group (HOPE or PAU), currently in jail, and employed. These were tested across sites, within HOPE, within PAU, and within each site. After Bonferroni corrections, none of the available variables was associated with consenting to participate, suggesting no systematic bias related to the likelihood of participation based on available data.

To assess whether HOPE and PAU who completed baseline interviews were similar at baseline, we fitted bivariate logistic regression models with the HOPE/PAU group indicator as the outcome and each other variable in the data set as a predictor. The p-values were stored and a Bonferroni correction applied. Any adjusted p-values less than 0.05 were examined. These steps were repeated for the full sample and each site. There were 229 variables in the baseline data, so we would expect at most 12 spurious significant predictors. After the Bonferroni correction, we found no significant differences between groups for the full sample and for the Massachusetts and Texas samples. One variable was a significant predictor for Arkansas and Texas, in both cases because of empty cells on the "select all that apply" response. Thus, our groups for whom we have baseline data appear to be equivalent.

To assess whether missingness at the 6- and 12-month interviews was associated with any baseline characteristics, we fitted bivariate logistic regression models with indicators of completion of 6-month and 12-month interviews as the outcome and each variable in the baseline data set as a predictor. Results were identical for the two sets of models, so we summarize the findings for the set of 12-month indicator models. The p-values from the set of logistic regression models were stored and a Bonferroni correction was applied. Any adjusted p-values less than 0.05 were further examined. These steps were

RESPONSE BIAS

Response bias analyses revealed no differences between the following groups: (1) those who completed a baseline instrument and those who did not; (2) HOPE and PAU baseline respondents; and (3) those who completed 6 or 12 month follow up interviews and those who did not (of those who completed baseline interviews).

followed for the full sample, the full PAU sample, the full HOPE sample, and for each site. There were 277 dependent variables, suggesting at most 14 spurious significant predictors. Only three variables were related to missingness at 12 months: (1) those who served in the Armed Forces were more likely to have completed a 12-month interview than those who had not served;[23] (3) those who agreed with the statement "Most successful people broke the law to get ahead in life." were more likely to have completed a 12-month interview than those who disagreed or were unsure;[24] and those who had higher scores on the question "Probation Officer would find out if I did not comply with conditions of my probation." were less likely to have completed a 12-month interview.[25] Thus, our groups who responded and did not respond to follow-up interviews appear to be largely similar and no further adjustments were made to the data.

As a result of these analyses, we have some confidence that the individuals who completed baseline interviews are represented of the full evaluation sample. Additionally, we can be confident that the follow-up is also representative.

Administrative Data

Administrative data were obtained for all 1,504 evaluation participants. The evaluation team worked with probation staff in the four DFE sites to understand the data maintained in their agencies' management information systems that could be used to describe HOPE and PAU probationers, and their probation experiences and outcomes. Through this data discovery process, the evaluation team documented the availability, quality, and format of the following administrative data in each site:

1. Demographics: date of birth, gender, race/ethnicity, marital status, number of minor children;
2. Education level and employment status;
3. Criminal history: previous probation and parole start/end dates, term length(s), conviction history (conviction dates, offenses, disposition), incarceration history (admission/release dates, offenses, prison infractions);
4. Current probation: convicted offense, risk score/level, other assessment scores (e.g., substance abuse), probation start date, length of probation term, conditions of probation, supervision contact types and dates, appointment compliance/noncompliance types and dates, restitution ordered and paid, probation fees paid;
5. Urine testing: initial color code and date assigned, date and types of color code changes, test dates, results, results contested, drug types tested;

[23] Only about 6% of those who completed a baseline interview reported having served in the Armed Services—63 of 977. 52% of those who served completed a 12-month interview compared with about 33% of those who had not served.

[24] 114 of the 945 who answered this question at baseline agreed with the statement compared with 651 who disagreed and 180 who were unsure. Of the 114, 46% completed a 12-month interview compared with about 32% of each of the other two groups.

[25] Response categories were 1 to 9 with 1 indicating "strongly disagree" and 9 indicating "strongly agree." The logistic parameter estimate for this variable was -0.077 (standard error = 0.032), Wald chi-square = 5.914, p = 0.015.

6. Violations of probation: date, violation type, warrant issue date, warrant served date, court hearing date, response to violation, date sanction imposed, sanction start/end dates, exceptions granted by judge and reason, sanction (jail) days ordered and served;
7. Treatment experiences: treatment referral dates, voluntary/mandatory status, treatment type (e.g., substance abuse, mental health), treatment setting (e.g., inpatient, outpatient), placement dates, treatment compliance and completion;
8. Recidivism (involvement in the criminal justice system after randomization): probation revocations (date of violation, reason, revocation date, length of term), convictions (date, offense, disposition), incarcerations (admission/release dates, offense, length of term).

Data availability varied by site so the evaluation team negotiated and executed site-specific data use agreements with the Arkansas Department of Community Corrections, Clackamas County Community Corrections, and Tarrant County Community Supervision and Corrections Department, and a research agreement with the Massachusetts Office of the Commissioner of Probation.

The evaluation team also requested access to the HOPE fidelity data that were collected by each site's HOPE program coordinator on behalf of Pepperdine University. These individual-level data, which were available only for those assigned to HOPE supervision, include:

1. Current probation: supervision type/level, warning hearing date
2. Urine testing: initial color code and date assigned, date and types of color code changes, test dates, results, results contested, drug types tested;
3. Violations of probation: violation date, violation type (e.g., drug test no show, late to drug test, abscond, new arrest), warrant issue date, warrant attempted service date, warrant served date, agency that served warrant, docket/court hearing date, docket event/court hearing type, response to violation, date sanction imposed, sanction start/end dates, exceptions granted by judge and reason, sanction (jail) days ordered and served (confinement date, release date);
4. Treatment experiences: treatment referral dates, treatment type (e.g., substance abuse), treatment setting (e.g., inpatient, outpatient), placement dates, treatment compliance and completion;
5. Recidivism (involvement in the criminal justice system after randomization): arrests for new charges (date of arrest, charge), probation revocations (date of violation, reason, revocation date, length of term), convictions (date, offense, disposition), incarcerations (admission/release dates, offense, length of term).

The evaluation team developed an administrative data transmission protocol that described procedures to ensure the secure delivery and storage of the requested data. The protocol was reviewed and approved by RTI's Institutional Review Board (IRB) in February 2014. The protocol was amended four times to accommodate site-level circumstances that required different procedures than those originally approved. All amendments were reviewed and approved by the IRB.

To perform the data match and extraction, the RTI evaluation team prepared and delivered to each probation agency and Pepperdine University a list of identifiers (e.g., name, date of birth, correctional ID number) for probationers randomized to HOPE or PAU. The original data acquisition schedule called for the transmission of an initial data extract from probation agencies in December 2013, followed by subsequent data extracts every 6 months through the end of the study period in March 2015. The

schedule was quickly abandoned due to difficulties accessing and preparing data. For example, the data use agreement with the Clackamas County (Oregon) probation agency and the research agreement with the Massachusetts probation agency were not fully executed until April and May 2014, respectively -- approximately 5 months after we initiated them.

Additionally, individual-level electronic probation data were not available in Massachusetts; we could obtain only electronic PDF documents of District Court probation case summaries and Superior Court probation officers' notes scanned to PDF documents by probation officers. These documents were provided one time after the end of the study period: We received District Court probation case summaries in September 2015, and Superior Court probation officers' case notes in February 2016. Evaluation team staff members reviewed these records and recorded relevant information into Excel spreadsheets. Data abstraction was completed in April 2016. These data had limitations particularly with respect to a lack of detail on drug testing. In some cases, the files simply indicated that the probationer was complying with testing requirements with no information on the number of times (and when) the individual had been tested or what drugs the individual was tested for.

Similar difficulties in accessing individual-level electronic data were encountered in Oregon and Texas. In Oregon, the evaluation research coordinator manually searched local data repositories and recorded relevant data into an Access database and Excel spreadsheets. Data entry was completed in December 2014. In Texas, information about probation violations (substantiated and unsubstantiated violations) and drug testing results were sourced from individual-level electronic data maintained by Tarrant County Community Supervision and Corrections Department. These data were merged with information about sanctions imposed in response to probation violations which were abstracted from hardcopy court orders and manually recorded by the evaluation research coordinator into an Excel spreadsheet. Data entry was completed in September 2015.

Administrative data acquisition from probation agencies was complete in March 2016 (Arkansas), February 2016 (Massachusetts), December 2014 (Oregon), and September 2015 (Texas). Pepperdine University provided two interim rounds of fidelity data (April and November 2014) and a final round in April 2015.

Because we were unable to access arrest data from the Federal Bureau of Investigation's National Crime Information Center, we developed alternative arrangements to obtain Criminal history data from state-level sources. This alternative approach limited our search for criminal history records to the four states with a DFE site:[26]

- In Arkansas, the Arkansas probation agency provided an extract of individual-level criminal history data (arrests) from the Arkansas Crime Information Center (ACIC) in July 2015.
- In Texas, the Tarrant County probation agency arranged access to criminal history records (arrests) through the Texas Crime Information Center (TCIC). The HOPE research coordinators attended training required to properly read TCIC criminal history records. The Tarrant County probation agency conducted the criminal history records search and

[26] We were unable to acquire data from the National Crime Information Center (NCIC), the proposed one-stop source for accessing nationwide criminal history data on all study subjects, due to a change in the Federal Bureau of Investigation's practice of releasing NCIC data for research purposes.

extraction (PDF files), and the HOPE research coordinator manually reviewed and recorded criminal history data into a spreadsheet. Data entry was completed in February 2016.

- After approving our research request, the Massachusetts Department of Criminal Justice Information Services provided an initial extract of individual-level electronic arraignment and court disposition data in April 2015. A subsequent search in July 2015 yielded data on an additional 50 probationers.
- In Oregon, state law states that the Oregon State Police is authorized to provide criminal records that contain any arrest less than 1-year old in which there has been no acquittal or dismissal and any records that contain a conviction. Given this limited access, the evaluation team turned to the Oregon Judicial Case Information Network (OJCIN), a subscription-based repository of historical court data (charges, convictions, sentences). In January 2015, the HOPE research coordinator began to manually search for and abstract criminal history data on study subjects into a spreadsheet. This effort continued through May 2015.

Data from all sources were de-identified, prepared (e.g., recoded, summarized), and combined for analysis.

Telephone-ACASI Component

A random sample of HOPE and PAU probationers who completed the ACASI baseline interviews were asked to participate in biweekly telephone mini-interviews. Enrollment in this part of the study began in June 2013 in Massachusetts, Oregon, and Texas and in September 2013 in Arkansas; enrollment ended in December 2013 in all four sites.

This part of the study was designed to examine whether attitudes of the HOPE participants changed over time compared to those on PAU.[27] There were nine question sets included in this component of the evaluation. These question sets were related to:

- Attitude toward supervision officer
- Deterrence related to drug use
- Deterrence related to compliance with conditions
- Perceived fairness of sanctions
- CSS-M attitudes toward the law subscale

[27] The T-ACASI study also attempted to assess the feasibility of this type of data collection. If many people who agree to participate phone in, this would be an economic method of obtaining information about changes in attitudes and behaviors over time. The consent rate was very high at over 90%, however this was subset from a group of probationers who had already consented to the ACASI interview conducted in the probation office. A total of 1,521 were randomly assigned to HOPE or PAU, of these 66% agreed to participate in the in-office baseline computer-assisted survey. A subset of this group was asked to participate in the phone-in mini-interviews. Most likely the consent rate is higher than it might be with another group of probationers who had not just agreed to participate in another survey. While the consent rate was satisfactory, the actual number of individuals who called in was low and diminished quickly over time. Only about half of the probationers who completed the baseline and consented to the T-ACASI protocol called at least once. The frequency of participants completing each module ranged from 1 (baseline) to 9. Our examination of the dropouts over time indicated an initial drop between baseline and the first phone-in. Call-ins also dropped after respondents called in the second time. Thus, additional retention efforts might focus on encouraging probationers to phone-in the first time and again after they called the second time.

- CSS-M tolerance for law violations
- CSS-M identification with criminal others
- Attitude toward judge
- Deterrence—likelihood, and severity of sanctions for rule violations
- Locus of control

The T-ACASI instrument is included in Appendix E. As all items included in the T-ACASI were given to participants at the baseline ACASI interview, the ACASI baseline was also the T-ACASI baseline interview.

The question sets are theoretically and practically important potential mediators of the effect of the HOPE program on outcomes such as recidivism and continued drug use. The constructs dealing with attitudes (e.g., the CSS-M subscales and the other attitude measures) are well established as key risk factors for recidivism, both new crimes and violations of supervision (Andrews and Bonta, 2003; Bucklen and Zajac, 2009). Anti-social attitudes influence how probationers process the deterrence message that is central to interventions such as HOPE. Similarly, constructs dealing with perceived fairness, locus of control, and deterrence related to drug use and compliance directly assess how probationers process the deterrence mechanism of HOPE (e.g., their beliefs about whether the sanctions they receive under HOPE are proportionate to their offense and their beliefs about the role that HOPE is playing in maintaining their sobriety). Finally, constructs related to housing, employment, and relationships represent key ecological stressors and indicators of social bonds that can influence how the probationer responds to HOPE (Sampson and Laub, 1995; Bucklen and Zajac, 2009; Lattimore and Visher, 2009; MacKenzie, 2002).

Individuals who agreed to participate were randomly assigned to one of seven 2-day weekly schedules (M/Th, Tu/F, W/Sa, Th/Su, F/M, Sa/Tu, or Su/W).[28] They were asked to call in on those two days to a 24-hour 1-800 telephone line. The T-ACASI system was programmed to provide one question set on each day. Individuals rotated through the question sets based on their assigned schedules. If an individual failed to call on his/her identified day, they could call in the following day. However, they could only call and respond to questions twice weekly and could only call in for 6 months (27 weeks). They provided responses using their telephone number pads. The question sets took no longer than 5 minutes to complete. All questions included "refused" as a valid response so that responding to questions was not required to obtain compensation.

As an incentive to participate, in three of the sites (Arkansas, Massachusetts, and Oregon) they were given rechargeable incentive cards to which we would periodically add funds based on their participation in the calls. In Texas, we provided probation fee credits. An experiment within this component was conducted in which we varied the amount and the timing of compensation. Specifically, compensation was offered as follows:

- Saline County, Arkansas: Those who agreed to participate were given a rechargeable gift card with an initial balance of $10; each time he or she called in and completed a mini-interview $2.50 was transferred to the card the following business day, for a total of up to $5 per week or $130 if they called twice a week for 6 months.
- Essex County, Massachusetts: Those who agreed to participate were given a rechargeable gift card with an initial balance of $20; each time he or she called in and completed a mini-

[28]The individual's study ID number contained a digit that indicated the schedule the individual was to adhere to.

interview $2.50 was transferred to the card the following business day, for a total of up to $5 per week or $140 if they called twice a week for 6 months.

- Clackamas County, Oregon: Those who agreed to participate were given a rechargeable gift card with an initial balance of $10; $20 was added to the card every four weeks if they participated at least once during that period, for a total of up to six payments and a total of $130.
- Tarrant County, Texas: Those who agreed to participate received a $20 probation fee credit; $20 in probation fee credits were provided every four weeks if they participated at least once during that period, for a total of up to six payments and a total of $140 paid toward their probation fees.

The opportunity to participate in the T-ACASI evaluation component was offered to 282 participants and 250 participants consented. Consent and participation rates are summarized in ***Exhibit 2-22***. Participation signified that the individual called into the call-in line at least once and the participation rate was calculated as the number completing at least one T-ACASI interview divided by the number who consented. The numbers vary across the sites due to differences in enrollment flow rates and, in Arkansas, because this component was in effect for a shorter period (beginning in September rather than June). On average, 90% of those who were offered the possibility of participating agreed to participate; however, of these only 49% called in at least once.

Exhibit 2-22. Summary of participation in T-ACASI experiment by site

Site	Offered	Consented	Consent Rate	Participated[1]	Participation Rate[1]
AR	29	29	100%	14	48.3%
MA	117	105	89.7%	56	53.3%
OR	88	74	84.1%	41	55.4%
TX	48	42	87.5%	16	38.1%
Overall	282	250	86.5%	127	50.8%
Average	--	--	90.3%	--	48.8%

[1] Called in at least once in addition to baseline ACASI data collection

Outcome Evaluation Analytic Methods

Data were available on all evaluation participants (N = 1504) from administrative sources and, thus, we examined outcomes for everyone without concern for selection or response bias. This was not the case for the interview data, where we had refusals and other cases where interviews were not conducted as was described above. As noted earlier, we did not identify any meaningful response bias with respect to the interview data.

Descriptive Analyses

We calculated means and standard deviations for all data overall by group, by site, and by group and site. Site-level means were compared using analysis of variance and between group differences were examined using t-tests. Full results are presented in tables in Appendices F, G, and H.

Survival Models

The primary measures of outcomes were the time to specific outcome events, including time to first arrest, time to first revocation, time to first arrest or revocation, time to first conviction, and time to first

residential treatment. Each outcome was examined visually using histograms and Kaplan-Meier plots. The latter were stratified by group (HOPE vs. PAU) overall and within each site. In our experience, the proportional hazards assumption rarely holds for criminal justice survival data, so parametric survival models were used instead of Cox survival models. To determine which distribution fit best, we fit intercept-only models using lognormal, exponential, Weibull, Gaussian, logistic, and log-logistic parametric survival models. The model with the smallest Akaike's Information Criterion (AIC; Akaike, 1974) was select as the best fitting model. ***For all recidivism outcomes, the lognormal distribution fit best.*** All models were fit with this distribution using R (R Core Team, 2016).

Where the data would support the analyses, we estimated sequential survival models in a gap analysis. For example, we examined the times between multiple violations with a series of models beginning with the time to the first violation, followed by time to the second violation conditioned on having a first violation, and so forth.

Competing Risk Survival Models

Competing hazard models were estimated to allow us to look at the relationship of HOPE program participation to first new arrest by offense type (person, drug, property, and public order/other). The combined cause-specific Cox competing risk survival model was used (Gerds & Scheike, 2015). Only HOPE versus PAU effects were examined for the competing risk models. Models were estimated for the full data set and for each site.

Count Models

Many of the outcomes of interest were best measured as counts. These count outcomes included recidivism charge count, recidivism arrest count, and recidivism conviction count. For each count outcome, several intercept-only models were fit and compared using Akaike's Information Criterion (AIC; Akaike, 1974). For each outcome, the model with the smallest AIC was taken as the best fitting model for that outcome and was used for all further multiple regression modeling of that outcome. Models tested included Gaussian, Tobit, Poisson, zero inflated Poisson, negative binomial, zero inflated negative binomial, and multinomial. These models were fit in R (R Core Team, 2016). ***The negative binomial model fit best for all outcomes.*** Negative binomial regression models with predictor variables (HOPE status, gender, race, etc.) were then estimated using PROC GENMOD in SAS (SAS Institute Inc., 2013). The models controlled for exposure time as an offset.

Multistate Models

Using all available data, a state history was constructed showing whether individuals were on the street, in jail, in residential treatment (RSAT), or revoked to prison, including durations in each state. The timing of arrests, violations, and deaths were also recorded. These histories provided detailed information on the variability among individual experiences of the probation experience. To visualize these data, the horizontal line plot (Tueller, Van Dorn, & Bobashev, 2016) as implement in the R (R Core Team, 2016) package longCatEDA (Tueller, 2016) was used. A horizontal line is produced for each participant. Color denotes which state they were in and the length of each colored segment denotes how long they were in that state, where the horizontal axis represents time. Participants' lines are stacked and sorted to reveal patterns in the probation histories. The plots were stratified by group and site to examine site and group differences. Events were superimposed on the lines using symbols (e.g., circles, x's, plus signs).

These multistate data were analyzed using several different analytic approaches. The simplest was the construction of transition matrices where transitions from the first to the second state, the second to the third state, and so on, were produced both numerically and graphically. This approach provided descriptive characterizations of transitions while ignoring the durations in each transition.

Competing risk models are extensions of survival models that examine the time (in days) to multiple competing events. For these transitions, the competing risks include being on the street, starting residential substance abuse treatment, violating terms of probation, being arrested, starting jail time, being revoked, or dying (which occurred too infrequently to be included in the analyses). We estimated the cause-specific Cox proportional hazard regression model for each site (Benichou and Gail, 1990). Since each probationer could cycle through multiple competing events, we extended the cause-specific Cox proportional hazard regression model to a gap analysis. This simply means that for each probationer's first, second, third, etc., event, a cause-specific Cox proportional hazard regression model was fit for the time since the last event until the current event. For simplicity, models for first 10 events were estimated though probationers experienced up to 77 events (or 76 transitions). A median of 9 transitions (10 events) were observed in the data. The only covariate was group membership, HOPE or PAU. The results were expressed as hazard ratios (HR), which are the risk of HOPE probationers making the transition to a given event relative to PAU probationers.

T-ACASI Analytic Approach

Descriptive and bivariate analyses were conducted to characterize the frequency of responding by study day, the numbers of times each survey was completed, and descriptive statistics for each outcome at baseline. Bivariate logistic regression, separately by site, was used to examine whether baseline values on each outcome, criminal history, and demographic variable predicted the likelihood to consent. These analyses inform whether modeling results might be biased due to (self) selection bias. T-tests were conducted to test for baseline group (HOPE vs. PAU) differences in each outcome. If randomization performed as expected, none of these t-tests would yield significant results. Bivariate logistic regression, separately by site, was used to examine whether baseline values on each outcome, criminal history, and demographic variable predicted completing one or more T-ACASI interview among cases consenting to the T-ACASI experiment. These analyses inform whether modeling results might be biased due to early dropout.

Mixed effects models were used to address repeated measures and unique numbers of observation within each probationer. Including the baseline ACASI interview, most scales were completed only two times, but individual scales were repeated up to nine times. Models were fit across sites (controlling for site and estimating all possible pairwise comparisons via a prior contrasts) and separately to responses within each site. Repeated measures under the mixed effects framework allow us to deal with these imbalanced data. A variance components covariance structure was used, which gave similar estimates to specifying a random-intercepts or an autoregressive covariance structure, but yielded a simpler model (i.e., fewer estimated parameters) than the alternatives.

Models were estimated using maximum likelihood. Predictors were time treated as a continuous predictor, HOPE/PAU group assignment, and the interaction between time and group assignment. Time was coded such that the time of the baseline ACASI interview was coded 0 and each subsequent call into the T-ACASI system was the elapsed time in days since the baseline interview. The parameter for time indicates the rate of change per day averaged over both groups. The parameter for group indicates the

magnitude of HOPE effect averaged over all time points, where PAU is the reference group. The parameter for the group by time interaction indicates the rate of change per day in the HOPE group above and beyond the overall rate of change per day.

2.4. Economic Evaluation

The economic evaluation addressed the following research questions:

1. What is the cost of starting and implementing HOPE?
2. What are the costs and (any) savings and how are these distributed among the agencies (level of government) participating in HOPE?
3. Is HOPE cost effective?

The analytic perspective of the economic analysis was the criminal justice system, meaning the analysis focused on costs and benefits of agencies such as the courts, probation, law enforcement, jail, and prison. Thus, we included the value of resources per individual associated with the following events:

- Intake
- Warning hearing
- Staffing meeting
- Office visit
- Drug test
- Violation hearing
- Arrest
- Jail days, including days in county jail, state jail, and county correctional centers (CCCs)
- Prison days
- Residential treatment days

We collected costs on each of these events and then combined the estimates to compute the net benefit of the HOPE program in terms of any differences in costs between HOPE probation and PAU probation.

ECONOMIC STUDY

Findings for the economic analyses are presented for three different subgroups: (1) 1,494 individuals for whom we had at least 6 months of data post study intake; (2) 1,291 individuals for whom we had at least 12 months of data; and (3) 625 individuals for whom we had at least 24 months of data.

Sample

For the economic evaluation, we created sub-samples from the larger sample of participants so that participants could be tracked for the same length of time. During the study, data on criminal justice events were collected until the end of the data collection period—which meant that about 3 years of data were available for those whose intake was early in the recruitment period and only a few months of data were available for participants whose intake was close to the end of the recruitment period. Although, the HOPE program was not formally time-limited, it was anticipated to last approximately 1 year. To make the best use of the available data, we identified three groups: the 625 probationers who had (at least) 24 months of post-intake data (311 in the PAU group and 314 in the HOPE group), the 1,291 probationers (649 PAU, 642

HOPE) who had 12 months of data after intake, and the 1,494 probationers (752 PAU, 742 HOPE), who had at least 6 months of post-intake data.

Data

We collected data on quantities and prices to estimate the costs of each treatment group. Quantities are the count of each of the criminal justice events (e.g., days in prison), and prices are the value per unit of event (e.g., cost per day in prison). ***Exhibit 2-23*** describes the quantity measures and their sources. Quantity data came from administrative and fidelity data obtained during the primary study, interviews with probation officers, and surveys with sites.

Exhibit 2-23. Measures and data sources of quantity of criminal justice events

Event	Measure of Event	Data Source
Intake	One per probationer	Study data
Warning hearings	One per probationer	Study data
Staffing meetings	Count of violations	Site surveys
Office visits	Count of office visits	Probation officer interviews
Drug tests	Count of drug tests	Study data
Violation hearings	Count of violations	Study data
Arrests	Count of arrests	Study data
State & county corrections	Count of days	Study data
Prison	Count of days	Study data
Residential treatment	Count of days	Study data

Price data came from a variety of sources. We calculated the price of a day in prison in each of the study states using data from the Bureau of Justice Statistics (BJS, 2012; BJS, 2015), and we conducted a literature and web scan to identify secondary sources for county jail costs in each state. We obtained the average price of a day in residential treatment from Shepard, Beaston-Blaakman, & Horgan (2003). Cohen (1988) estimated the weighted average of arrest costs by separated cost components; we used the sum of the Cohen's investigation, arrest, and booking components as the average price of arrest in this study. We obtained the price of drug testing materials from probation officers during phone interviews. To account for personnel costs, we collected wage rates for the Bureau of Labor Statistics (BLS, 2014) for criminal justice personnel associated with each event. We combined the wage rates with the time and attendance information from probation officer interviews and site surveys to calculate a per-unit price for each event. We converted all prices to 2015 dollars using the Consumer Price Index. ***Exhibit 2-24*** shows the price measures and their sources.

Exhibit 2-24. Measures and data sources of price per criminal justice event

Event	Unit	Data Sources
Intake	per intake	BLS, 2014; PO interviews
Warning hearings	per probationer per hearing	BLS, 2014; Site surveys
Staffing meetings	per probationer per meeting	BLS, 2014; Site surveys
Office visits	per office visit	BLS, 2014; PO interviews
Drug tests	per drug test	BLS, 2014; PO interviews
Violation hearings	per probationer per hearing	BLS, 2014; Site surveys
Arrests	per arrest	Cohen, 1988
State & county corrections	per inmate per day	Association of Arkansas Counties, 2011; Norman, 2012; Baker et al., 2015; Essex Sheriff's Department, 2015; Nice, 2002; Yáñez-Correa & Totman, 2010
Prison	per inmate per day	BJS, 2012; BJS, 2015
Residential treatment	per client per day	Shepard et al., 2003

Analysis

We used a net benefit approach to calculate the costs of HOPE and PAU. The per-probationer monthly cost C_{it} was calculated as the sum of the product of the prices P_{ite} and quantities Q_{ite} for each of the ten criminal justice events described above:

$$C_{it} = \sum_{e=1}^{10} P_{ite}\, Q_{ite}$$

for $i = 1, \ldots, n$ probationers, $t = 1, \ldots, 24$ months, and $e = 1, \ldots, 10$ criminal justice events. We then compared average monthly and average total costs of HOPE probationers to PAU probationers. All calculations were performed using Stata MP 14.

The cost calculations have four main assumptions:

1. County correctional centers (CCCs) and state jails had the same cost per day as jail. We made this assumption because there are no reliable estimates of the costs of CCC's.
2. Probation officer visits were estimated based on site-level information from probation officers about the average number of office visits per month for HOPE and PAU probationers. The average was assigned to each probationer in each month, adjusted for any time not in the community. Specifically, each probationer's assigned number of monthly office visits was weighted by the number of days enrolled in the study and in the community (not in jail, prison, or residential treatment) as a proportion of the total number of days.
3. Probationers in a site with the same warning hearing date were assumed to attend the same warning hearing. Study fidelity data reported warning hearing dates for each HOPE probationer but did not indicate whether probationers had different hearings on the same day. We assigned an equal share of the warning hearing cost to each probationer at each

warning hearing. We imputed the site-level average probationer warning hearing cost to three HOPE probationers without warning hearing dates in the administrative data.

4. Staff meeting costs were adjusted for Arkansas. Oregon and Texas reported staffing meeting time and attendance on a per violation basis, while Arkansas reported it on a per month basis. We converted Arkansas' staffing meeting cost to a per violation basis by summing the total number of violations per month and applying cost per violation per month proportionately to each probationer violation

3. Evaluation Participants

A total of 1,580 individuals were randomly assigned to HOPE (794) or PAU (786) (***Exhibit 3-1***) and 1,504 individuals comprised the final study-eligible sample (743 HOPE and 761 PAU). The remaining 76 individuals (4.8%) were study ineligible, including 68 individuals (4.3%) who were deemed program ineligible, and 8 individuals (0.5%) who were randomized twice.

Exhibit 3-1. Subjects randomized to the HOPE DFE

	HOPE		PAU		ALL CASES	
	N	% of All Cases	N	% of All Cases	N	% of All Cases
TOTAL ALL CASES	794	50.2	786	49.8	1,580	100.0
	N	**% of All HOPE**	**N**	**% of All PAU**	**N**	**% of All Cases**
Study Eligible	743	93.6	761	96.8	1,504	95.2
Study Ineligible	51	6.4	25	3.2	76	4.8
Program ineligible	45	5.7	23	2.9	68	4.3
Already randomized	6	0.7	2	0.3	8	0.5

In the Arkansas site, 361 cases were randomly assigned to HOPE (191) or PAU (170) between August 2012 and September 2014 (***Exhibit 3-2***). A total of 342 individuals (95%) comprised the final study eligible sample. The remaining 19 individuals were study ineligible, including 15 individuals who were deemed program ineligible, and 4 individuals who were randomized twice.

Exhibit 3-2. Subjects randomized to the HOPE DFE, Benton County, AR

	HOPE		PAU		ALL CASES	
	N	% of All Cases	N	% of All Cases	N	% of All Cases
TOTAL ALL CASES	191	52.9	170	47.1	361	100.0
	N	**% of All HOPE**	**N**	**% of All PAU**	**N**	**% of All Cases**
Study Eligible	179	93.7	163	95.9	342	94.7
Study Ineligible	12	6.3	7	4.1	19	5.3
Program ineligible	9	4.7	6	3.5	15	4.2
Already randomized	3	1.6	1	0.6	4	1.1

In the Massachusetts site, 423 cases were randomly assigned to HOPE (209) or PAU (214) between October 2012 and July 2014 (***Exhibit 3-3***). A total of 392 individuals comprised the final study eligible sample and 31 individuals were determined to be study ineligible (29 who were program ineligible, and 2 who were randomized twice).

Exhibit 3-3. Disposition of subjects randomized to the DFE, Essex County, MA

	HOPE		PAU		ALL CASES	
	N	% of All Cases	N	% of All Cases	N	% of All Cases
TOTAL ALL CASES	209	49.4	214	50.6	423	100.0
	N	% of All HOPE	N	% of All PAU	N	% of All Cases
Study Eligible	189	90.4	203	94.9	392	92.7
Study Ineligible	20	9.6	11	5.1	31	7.3
Program ineligible	18	8.6	11	5.1	29	6.8
Already randomized	2	1.0	0	0.0	2	0.5

In the Oregon site, 412 cases were randomly assigned to HOPE (202) or PAU (210) between August 2012 and December 2013 (***Exhibit 3-4***). A total of 394 individuals comprised the final study eligible sample. The remaining 18 individuals were study ineligible because they were program ineligible.

Exhibit 3-4. Disposition of subjects randomized to the DFE, Clackamas County, OR

	HOPE		PAU		ALL CASES	
	N	% of All Cases	N	% of All Cases	N	% of All Cases
TOTAL ALL CASES	202	49.0	210	51.0	412	100.0
	N	% of All HOPE	N	% of All PAU	N	% of All Cases
Study Eligible	190	94.1	204	97.1	394	95.6
Study Ineligible	12	5.9	6	2.9	18	4.4
Program ineligible	12	5.9	6	2.9	18	4.4
Already randomized	0	0.0	0	0.0	0	0.0

In the Texas site, 384 cases were randomly assigned to HOPE or PAU (192 in each group) between August 2012 and September 2014 (***Exhibit 3-5***). A total of 376 individuals (97.9%) comprised the final study eligible sample. The remaining 8 individuals were study ineligible, including 6 individuals who were program ineligible, and 2 individuals who were randomized twice.

Exhibit 3-5. Disposition of subjects randomized to the DFE, Tarrant County, TX

	HOPE		PAU		ALL CASES	
	N	% of All Cases	N	% of All Cases	N	% of All Cases
TOTAL ALL CASES	192	50.0	192	50.0	384	100.0
	N	% of All HOPE	N	% of All PAU	N	% of All Cases
Study Eligible	185	96.4	191	99.5	376	97.9
Study Ineligible	7	3.6	1	0.5	8	2.1
Program ineligible	6	3.1	0	0.0	6	1.6
Already randomized	1	0.5	1	0.5	2	0.5

Exhibit 3-6 summarizes the final distribution of eligible study subjects randomized to the DFE overall and by site. At the site level, distribution of cases between HOPE and PAU mirrored the distribution of cases overall.

Exhibit 3-6. Subjects enrolled in the DFE by site and group

Site	HOPE N	HOPE %	PAU N	PAU %	Total N	% Total
Saline County, AR	179	52.3	163	47.7	342	22.7
Essex County, MA	189	48.2	203	51.8	392	26.1
Clackamas County, OR	190	48.2	204	51.8	394	26.2
Tarrant County, TX	185	49.0	191	51.0	376	25.0
Total	743	49.4	761	50.6	1,504	100.0

3.1. Baseline Demographic Characteristics of DFE Subjects

The baseline characteristics of the full study sample were identified from administrative data sources. HOPE and PAU study participants were similar in their pre-study characteristics (***Exhibit 3-7***). They were, on average, about 31 years old, male, white, and high risk, with 7 prior arrests and 3+ prior convictions. Most had a history of arrest for a variety of offenses. Individuals in the study were on probation for a drug (31%), property (30%), person (24%), or public order/other (15%) offense. The PAU group was significantly more likely than the HOPE group to have a person charge (26% versus 21%) and less likely to have a public order/other charge (13% versus 18%).

Exhibit 3-7. Characteristics of subjects enrolled in the DFE, overall and by group

Characteristics	Overall % (SD)	HOPE% (SD)	PAU% (SD)
Age at intake (mean)* **	31.0 (10.37)	30.6 (10.06)	31.5 (10.66)
Male = 1 * **	0.81 (0.40)	0.81 (0.39)	0.80 (0.40)
Race			
White* **	0.69 (0.46)	0.69 (0.46)	0.68 (0.47)
Black* **	0.16 (0.37)	0.16 (0.37)	0.16 (0.37)
Hispanic* **	0.13 (0.34)	0.13 (0.33)	0.13 (0.34)
Other* **	0.02 (0.15)	0.02 (0.14)	0.02 (0.15)
Age at first arrest (mean)* **	22.1 (7.78)	21.9 (7.64)	22.4 (7.91)
Number of prior arrests (mean)* **	7.3 (8.13)	7.4 (8.46)	7.3 (7.82)
Has a prior person charge* **	0.56 (0.50)	0.57 (0.50)	0.56 (0.50)
Has a prior property charge* **	0.74 (0.44)	0.75 (0.43)	0.74 (0.44)
Has a prior drug charge* **	0.66 (0.48)	0.66 (0.47)	0.65 (0.48)
Has a prior public order/other charge* **	0.77 (0.42)	0.75 (0.43)	0.79 (0.41)
Number of prior convictions (mean)* **	3.5 (4.42)	3.6 (4.80)	3.4 (4.01)
High Risk* **	0.55 (0.50)	0.55 (0.50)	0.54 (0.50)
Study Offense			
Person* **	0.24 (0.42)	**0.21† (0.41)**	**0.26† (0.44)**
Property* **	0.30 (0.46)	0.31 (0.46)	0.29 (0.46)
Drug* **	0.31 (0.46)	0.30 (0.46)	0.32 (0.47)
Public Order/Other* **	0.15 (0.36)	**0.18‡ (0.38)**	**0.13 ‡ (0.33)**
N	1504	743	761

***Subject characteristics differed across sites ($p < 0.001$).

†HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.

In Arkansas, there were no significant differences between those in HOPE and PAU. Those enrolled in the study were about 32 years of age (somewhat older than the full study sample), male, and not assessed as high risk (***Exhibit 3-8***). Unlike the full study sample (and the other three sites), few of the study enrollees were high risk (3%). Most were assessed as low risk (73% of the HOPE and 83% of PAU groups; significantly different at $p < 0.02$). Although unable to conclusively verify, the most likely explanation appears to be that low-risk probationers who violated their conditions became HOPE-eligible and were randomly assigned to either HOPE or PAU without conducting (or, perhaps, recording) a new risk assessment that would have resulted in a reclassification to high (or medium) risk. Arkansas probationers had less prior criminal justice system involvement than the overall study sample with about 4 prior arrests and about 2 prior convictions. The offense charge for their current probation was for a property offense (37%), drug offense (24%), person offense (23%), and public order/other offense (16%). There were no differences in current offense charge between the HOPE and PAU groups.

Exhibit 3-8. Characteristics of subjects enrolled in the DFE, Saline County, AR

Characteristics	Overall % (SD)	HOPE % (SD)	PAU % (SD)
Age at intake (mean)	32.3 (10.21)	32.1 (9.73)	32.5 (10.73)
Male = 1	0.73 (0.44)	0.77 (0.42	0.69 (0.46)
Race			
White	0.85 (0.36)	0.87 (0.34)	0.83 (0.38)
Black	0.14 (0.35)	0.12 (0.33)	0.16 (0.37)
Hispanic	0.01 (0.08)	0.01 (0.07)	0.01 (0.08)
Other	0.01 (0.08)	0.01 (0.07)	0.01 (0.08)
Age at first arrest (mean)	27.2 (9.63)	26.9 (9.46)	27.5 (9.83)
Number of prior arrests (mean)	4.4 (3.38)	4.3 (3.27)	4.5 (3.51)
Has a prior person charge	0.52 (0.50)	0.54 (0.50)	0.49 (0.50)
Has a prior property charge	0.69 (0.46)	0.70 (0.46)	0.68 (0.47)
Has a prior drug charge	0.59 (0.49)	0.57 (0.50)	0.61 (0.49)
Has a prior public order/other charge	0.74 (0.44)	0.73 (0.45)	0.75 (0.43)
Number of prior convictions (mean)	1.7 (1.13)	1.8 (1.21)	1.7 (1.04)
High Risk	0.03 (0.17)	0.03 (0.17)	0.03 (0.17)
Study Offense			
Person	0.23 (0.42)	0.22 (0.41)	0.25 (0.44)
Property	0.37 (0.48)	0.34 (0.48)	0.40 (0.49)
Drug	0.24 (0.43)	0.25 (0.44)	0.23 (0.42)
Public Order/Other	0.16 (0.36)	0.19 (0.39)	0.12 (0.32)
N	342	179	163

NOTE: There were no differences in baseline characteristics between those assigned to HOPE and those assigned to PAU.

In Massachusetts, there were no significant differences in pre-study characteristics between the HOPE and PAU study groups (***Exhibit 3-9***). The Massachusetts study participants were somewhat older on average than the overall study population (nearly 34 years old compared to the overall average of 31), more likely to be male (88% versus 81%), and more likely to be high risk (72% versus 55%). Among study participants across the four DFE sites, the Massachusetts study participants had the most extensive criminal histories, typically experiencing their first arrest at 20, and accumulating an average of 13 prior arrests and nearly 6 prior convictions. The Massachusetts participants were more likely than those from the other sites to have a current charge for a person offense (50% versus 24% for the study population overall); the current offenses of other Massachusetts participants were property (22%), public order/other (17%), and drug (11%). The more serious criminal justice involvement of the Massachusetts' sample likely reflects the more serious involvement of the Superior Court probationers—who were serving probation following release from prison.

Exhibit 3-9. Characteristics of subjects enrolled in the DFE, Essex County, MA

Characteristics	Overall % (SD)	HOPE % (SD)	PAU % (SD)
Age at intake (mean)	33.7 (11.14)	33.7 (10.75)	33.6 (11.51)
Male = 1	0.88 (0.32)	0.88 (0.33)	0.89 (0.32)
Race			
White	0.68 (0.47)	0.69 (0.46)	0.67 (0.47)
Black	0.08 (0.26)	0.08 (0.27)	0.07 (0.26)
Hispanic	0.18 (0.39)	0.18 (0.39)	0.18 (0.39)
Other	0.06 (0.24)	0.05 (0.22)	0.08 (0.27)
Age at first arrest (mean)	20.0 (6.53)	20.0 (6.65)	19.9 (6.44)
Number of prior arrests (mean)	13.0 (11.87)	13.5 (12.51)	12.5 (11.25)
Has a prior person charge	0.86 (0.35)	0.86 (0.35)	0.85 (0.36)
Has a prior property charge	0.81 (0.39)	0.80 (0.40)	0.82 (0.38)
Has a prior drug charge	0.57 (0.50)	0.54 (0.50)	0.60 (0.49)
Has a prior public order/other charge	0.93 (0.26)	0.92 (0.27)	0.93 (0.25)
Number of prior convictions (mean)	5.8 (6.45)	6.4 (7.41)	5.2 (5.38)
High Risk	0.72 (0.45)	0.74 (0.44)	0.70 (0.46)
Study Offense			
Person	0.50 (0.50)	0.46 (0.50)	0.53 (0.50)
Property	0.22 (0.42)	0.25 (0.43)	0.20 (0.40)
Drug	0.11 (0.31)	0.10 (0.30)	0.12 (0.33)
Public Order/Other	0.17 (0.38)	0.19 (0.39)	0.15 (0.36)
N	392	189	203

NOTE: There were no differences in pre-study characteristics between those assigned to HOPE and those assigned to PAU.

Oregon subjects were about the same age as the study population overall (31 years), 88% were white, and 83% were male (***Exhibit 3-10***). Among the study groups in the four sites, the Oregon study group had the highest proportion of high-risk probationers (88% compared with 72% in Massachusetts, 52% in Texas, and 3% in Arkansas). On average, the Oregon study participants had fewer prior arrests than the overall study population (6.0 versus 7.3) but a greater number of prior convictions (4.3 versus to 3.5). ***There were a few baseline differences between those assigned to the two groups.*** Those assigned to HOPE were significantly more likely than those assigned to PAU to be assessed as high risk (93% versus 84%). The HOPE probationers were more likely than the PAU probationers to have a prior drug charge (80% versus 67%), while study participants assigned to PAU were significantly more likely than those assigned to HOPE to have a prior person charge (53% versus 42%) and a prior public order/other charge (81% versus 71%). There were also significant differences between the two groups in current offense. Although both groups were equally likely to have a current offense as a drug offense (43% HOPE, 46% PAU), HOPE probationers were more likely than PAU to have a current public order/other charge (25% HOPE, 16% PAU) and less likely to have a person offense charge (15% HOPE, 24% PAU). The two groups had similar percentages with a property offense (17% HOPE, 15% PAU).

Exhibit 3-10. Characteristics of subjects enrolled in the DFE, Clackamas County, OR

Characteristics	Overall % (SD)	HOPE % (SD)	PAU % (SD)
Age at intake (mean)	30.8 (9.91)	30.0 (9.89)	31.5 (9.91)
Male = 1	0.83 (0.38)	0.83 (0.38)	0.83 (0.38)
Race			
White	0.88 (0.33)	0.87 (0.33)	0.88 (0.33)
Black	0.05 (0.22)	0.05 (0.22)	0.05 (0.22)
Hispanic	0.06 (0.23)	0.05 (0.22)	0.06 (0.24)
Other	0.02 (0.13)	0.02 (0.14)	0.02 (0.12)
Age at first arrest (mean)	22.9 (6.17)	22.4 (5.52)	23.3 (6.70)
Number of prior arrests (mean)	6.0 (6.05)	5.8 (5.93)	6.2 (6.17)
Has a prior person charge	0.48 (0.50)	**0.42† (0.49)**	**0.53† (0.50)**
Has a prior property charge	0.69 (0.46)	0.73 (0.45)	0.66 (0.48)
Has a prior drug charge	0.73 (0.440	**0.80‡ (0.40)**	**0.67‡ (0.47)**
Has a prior public order/other charge	0.76 (0.43)	**0.71† (0.46)**	**0.81† (0.39)**
Number of prior convictions (mean)	4.3 (3.91)	4.1 (3.78)	4.4 (4.04)
High Risk	0.88 (0.33)	**0.93‡ (0.26)**	**0.84‡ (0.37)**
Study Offense			
Person	0.20 (0.40)	0.15† (0.36)	0.24† (0.43)
Property	0.16 (0.36)	0.17 (0.38)	0.15 (0.36)
Drug	0.44 (0.50)	0.43 (0.50)	0.46 (0.50)
Public Order/Other	0.20 (0.40)	**0.25† (0.44)**	**0.16† (0.36)**
N	394	190	204

†HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.

The Texas study group was the youngest, on average—nearly 28 years old compared with about 31 in Oregon, 32 in Arkansas, and 34 in Massachusetts (***Exhibit 3-11***). ***The HOPE probationers were younger, on average, than the PAU group—26.5 years compared with 28.4 years.*** There were fewer male probationers in Texas than the study population overall (77% compared with 81%), although there was no significant difference between the two Texas groups. Additionally, there were fewer white probationers in the Texas study group than the study population overall (35% compared with 69%). Compared with the average participant in the overall study group, the average Texas study participant experienced his/her first arrest at a younger age (19 versus 22), had fewer prior arrests (5.4 versus 7.3), and fewer prior convictions (2.1 versus 3.5). The Texas study group was slightly less likely than the overall study group to be assessed as high risk (52% versus 55%). Among the Texas study participants, those assigned to HOPE were significantly more likely than those assigned to PAU to have a prior person charge (44% versus 32%; significant $p < 0.05$). The two groups were similar on the likelihood of having a prior drug charge, property charge, and public order/other charge. Current offenses were similar for the two groups. About 48% had a current property offense; 44% had a current drug offense; 8% had a current public order/other offense; and 1% had a current person offense.

Exhibit 3-11. Characteristics of subjects enrolled in the DFE, Tarrant County, TX

Characteristics	Overall % (SD)	HOPE % (SD)	PAU % (SD)
Age at intake (mean)	27.5 (9.10)	**26.5† (8.26)**	**28.4† (9.78)**
Male = 1	0.77 (0.42)	0.77 (0.42)	0.76 (0.43)
Race			
White	0.35 (0.48)	0.33 (0.47)	0.37 (0.48)
Black	0.38 (0.49)	0.39 (0.49)	0.37 (0.48)
Hispanic	0.27 (0.45)	0.28 (0.45)	0.27 (0.44)
Other	0.01 (0.05)	0.01 (0.07)	0.00 (0.00)
Age at first arrest (mean)	19.0 (5.95)	18.5 (5.67)	19.5 (6.17)
Number of prior arrests (mean)	5.4 (4.70)	5.6 (5.30)	5.3 (4.04)
Has a prior person charge	0.38 (0.49)	**0.44† (0.50)**	**0.32† (0.47)**
Has a prior property charge	0.77 (0.42)	0.76 (0.43)	0.78 (0.42)
Has a prior drug charge	0.73 (0.44)	0.72 (0.45)	0.74 (0.44)
Has a prior public order/other charge	0.65 (0.48)	0.66 (0.48)	0.64 (0.48)
Number of prior convictions (mean)	2.1 (2.51)	2.2 (2.62)	2.0 (2.41)
High Risk	0.52 (0.50)	0.53 (0.50)	0.51 (0.50)
Study Offense			
Person	0.01 (0.10)	0.02 (0.13)	0.01 (0.07)
Property	0.48 (0.50)	0.51 (0.50)	0.46 (0.50)
Drug	0.44 (0.50)	0.40 (0.49)	0.47 (0.50)
Public Order/Other	0.08 (0.26)	0.08 (0.27)	0.07 (0.25)
N	376	185	191

†HOPE and PAU differ at $p < 0.05$.

3.2. Interview Results

Among 1,504 probationers in the study sample, 978 (65%) completed a Wave 1 interview (***Exhibit 3-12***). There was variation by site and wave. Overall, response rates were 65% at baseline, 35% at 6-month follow-up, and 35% at 12-month follow-up. Because we attempted to obtain interviews from everyone at each wave, we obtained at least one interview from 1148 (76%) of the 1504 evaluation-eligible participants and at least one follow-up interview for 711 (47%). As noted in Section 2, tests for response bias found no differences between those interviewed and those not interviewed—overall at baseline and among baseline respondents at follow-up.

Exhibit 3-12. HOPE interview completion summary counts

Site	AR	OR	TX	MA	Total
Baseline Intake					
Program Ineligible	19	18	8	31	76
Non-interview	85	50	76	9	220
Refusals	87	99	83	37	306
Completes	170	245	217	346	978
Baseline Intake Totals (without ineligibles)	342	394	376	392	1504
Baseline Response Rates (%)	**50**	**62**	**58**	**88**	**65**
6-month Intake					
Fielded	342	394	376	392	1504
Non-interview	184	249	260	204	897
Refusals	13	17	25	16	71
Completes	145	128	91	172	536
6-month Response Rates (%)	**42**	**32**	**24**	**44**	**36**
12-month Intake					
Fielded	256	394	333	329	1312
Non-interview	108	269	216	201	794
Refusals	11	9	29	10	59
Completes	137	116	88	118	459
12-month Response Rates (%)	**54**	**29**	**26**	**36**	**35**

3.3. Interview Findings by Wave

This section describes the results of the three waves of HOPE and PAU, including demographic characteristics, education and employment, homelessness, mental health, service needs and receipt, criminal justice involvement, and family and peers. For brevity, the term "probationers" will be used to refer to interview participants (e.g., HOPE probationers). Detailed data tables containing means, standard deviations, t-statistics, and p-levels are included in Appendix F.

This section also presents information about the characteristics of interview participants at the site level and at the site level by study group. In general, only differences between sites and between study groups at the site-level level are discussed. For more information, site-level data tables containing means, standard deviations, F-statistics, and p-levels are provided in Appendix G, and detailed data tables containing means, standard deviations, t-statistics, and p-levels for study group comparisons at the site-level are provided in Appendix H.

RESPONSE BIAS

Response bias analyses suggested no differences between those interviewed and those not interviewed (see Section 2.3).

Demographic Characteristics

Exhibit 3-13 shows the self-reported demographic characteristics of HOPE and PAU probationers at each interview wave. Results were similar across waves. At each interview wave, about 80% of HOPE and PAU probationers were male, nearly 70% were white, and about 15% were black. In Wave 1, about 15% of HOPE and PAU probationers were of Hispanic origin, a slightly higher proportion than in Wave 2 or Wave 3 although the difference was not statically significant. The only significant between-group differences were age and service in the Armed Forces. On average, HOPE probationers were younger than PAU probationers at Wave 1 (30.6 years versus 32 years). At the Wave 1 interview, more HOPE probationers than PAU probationers reported having ever served in the Armed Forces—although the percentages were small (8% versus 5%).

Exhibit 3-13. Demographic characteristics of interview participants, by group and wave

	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
Characteristic	HOPE	PAU	HOPE	PAU	HOPE	PAU
Age at baseline interview (mean)	**30.6***	**32.0**	32.1	32.4	32.2	33.5
Male	81.9	79.7	82.5	80.5	79.3	80.7
White	66.1	68.5	69.0	68.1	66.9	70.8
Black	14.7	15.1	14.5	16.2	14.7	15.1
American Indian/Alaska Native	2.5	1.2	1.7	1.3	1.1	0.5
Native Hawaiian/Pacific Islander	0.1	0.2	0.3	0.0	0.4	0.0
Asian	0.0	0.2	0.0	0.4	0.0	0.5
More than one race	5.9	4.9	5.6	7.9	7.1	6.8
Other race	9.8	9.9	8.9	6.1	9.8	6.3
Hispanic origin	15.0	15.1	13.8	13.4	13.1	11.5
Born in the US	95.3	97.1	98.4	97.0	97.8	95.8
Born in Mexico	1.0	0.6	0.7	0.0	0.7	1.0
Born in another country	3.7	2.3	1.0	3.0	1.5	3.1
Speak mainly English	94.3	96.1	96.4	94.4	96.6	96.9
Speak mainly Spanish	5.3	3.5	3.6	5.2	2.6	2.1
Speak mainly another language	0.4	0.4	0.0	0.4	0.7	1.0
Ever served in the Armed Forces	**8.0***	**4.7**	11.8	8.7	12.4	8.9

*HOPE and PAU differ at $p < 0.05$.

Demographic characteristics of probationers varied by site (see tables in Appendix G). For example, at Wave 1, the average age ranged from 28.4 years old in Texas to 33.5 years old in Massachusetts. The proportion of Wave 1 probationers who were male ranged from 72.9% in Arkansas to 87.0% in Massachusetts. Race, Hispanic origin, and language also varied by site. For example, the Oregon study group had the highest proportion of white probationers and Texas had the lowest rate (79.5% and 39.5%, respectively). Black probationers comprised 40.0% of the Texas study group, while only 3.3% of the probationers in the Oregon study group were black. Texas had the highest rate of probationers of

Hispanic origin and Arkansas had the lowest rate (26.5% and 1.2%, respectively). The highest rate of English speakers was in Arkansas and the lowest rate was in Massachusetts (100% and 92.5%, respectively). There was no difference in nativity or service in the Armed Forces among the sites.

There were few differences in demographic characteristics between HOPE and PAU probationers at the site level (Appendix H). At the Wave 1 interview, HOPE probationers were younger than PAU probationers in Texas (26.7 years versus 30.2 years). With respect to race, 3.9% of HOPE probationers in Arkansas identified as American Indian or Alaska Native; none of their PAU counterparts identified themselves as such. With respect to nativity, 3.6% of HOPE probationers in Texas reported that they were born outside of the United States or Mexico; no PAU probationers in Texas reported the same. In Texas, there was also a difference among study groups with respect to language. HOPE probationers were more likely than PAU probationers to report that they speak mainly Spanish (9.9% versus 2.8%) Finally, HOPE probationers in Massachusetts were more likely than PAU probationers to report having served in the Armed Forces (10.4% versus 4.4%).

At the Wave 2 and Wave 3 interviews, there were few differences between HOPE and PAU probationers than at the Wave 1 interview. At the Wave 2 interview, 14.6% of PAU probationers in Massachusetts identified as being more than one race, compared to 3.4% of HOPE probationers. At the Wave 3 interview, there were differences between the study groups with respect to race in Arkansas and nativity in Massachusetts. In Arkansas, 25.0% of HOPE probationers identified as black, compared to 8.6% of their HOPE counterparts. Additionally, 85.2% of HOPE probationers compared to 69.6% of PAU probationers identified as white. In Massachusetts, all HOPE probationers reported being born in the United States, compared to 93.3% of PAU probationers.

Education and Employment

Exhibit 3-14 shows education status among probationers at each interview wave. About 50% of HOPE probationers and 54.5% of PAU probationers reported attaining a high school diploma or GED and nearly 20% of HOPE and PAU probationers reported having attend a vocational school or college without getting a certificate or degree. At the Wave 1 interview, 8% of HOPE probationers and 9% of PAU probationers reported being in school. Fewer HOPE probationers than PAU probationers reported being in school at Wave 2 (5% versus 10%) and Wave 3 (6% versus 13%).

Exhibit 3-14. Self-reported education status, by group and wave

	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
Education	**HOPE**	**PAU**	**HOPE**	**PAU**	**HOPE**	**PAU**
Currently in school	8.4	9.1	**5.3***	**10.8**	**6.4***	**13.0**
Highest level of education						
College degree	5.1	4.3	5.3	5.2	4.5	7.3
Vocational school or college, no degree	18.2	19.3	17.4	18.6	19.5	19.8
High school diploma or GED	49.7	54.5	52.6	56.7	51.5	52.1
Less than high school	24.1	20.3	22.4*	15.6	20.7	18.8
None	2.9	1.7	2.3	3.9	3.8	2.1

* HOPE and PAU differ at $p < 0.05$.

There were several differences among the sites in probationer education status (Appendix G). ***At Wave 1, the rate of current school enrollment varied across the sites, ranging from 5% of probationers in Oregon to 14% in Texas.*** At the Wave 2 interview, the highest level of educational attainment varied by site. Specifically, 1% of Arkansas probationers, 2% of Massachusetts probationers, 3% of Oregon probationers, and 8% of Texas probationers reported that they had no education. There were no differences in education status among the sites at the Wave 3 interview.

At the site level, there were a few differences between HOPE and PAU probationers with respect to education (Appendix H). At Wave 1, more HOPE probationers than PAU probationers in Arkansas reported that they had no education (4% versus 0%). In Massachusetts, PAU probationers were more likely than HOPE probationers to report that they had attained a high school diploma or GED (59% versus 48%) and less likely to report that they had no education (0.6% versus 4%, respectively). At Wave 2, PAU probationers in Texas were more likely than HOPE probationers to report they were currently in school (20% versus 6%) and had attained a high school diploma or GED (64% versus 40%). HOPE probationers in Texas were more likely than PAU probationers to report experience with vocational school or college (23% versus 8%). In Oregon, HOPE probationers were more likely than PAU probationers to report that they had less than a high school education at Wave 1 (32% versus 20%), Wave 2 interview (33% versus 17%), and Wave 3 interview (34% versus 16%). Additionally, PAU probationers in Oregon were more likely than HOPE probationers to report that they had attained a high school diploma or GED at the Wave 3 interview (84% versus 64%). There were no other differences between study groups with respect to education at the Wave 3 interview.

At each interview wave, probationers were asked about their current or most recent employment, wages, hours, and benefits. At the Wave 1 interview, HOPE and PAU probationers reported similar rates of employment (34% and 31%; ***Exhibit 3-15***). Wave 2 and Wave 3 employment rates were higher than Wave 1 rates for both groups. Employment rates for HOPE probationers at Wave 2 and Wave 3 were higher than but not significantly different from those for PAU probationers.

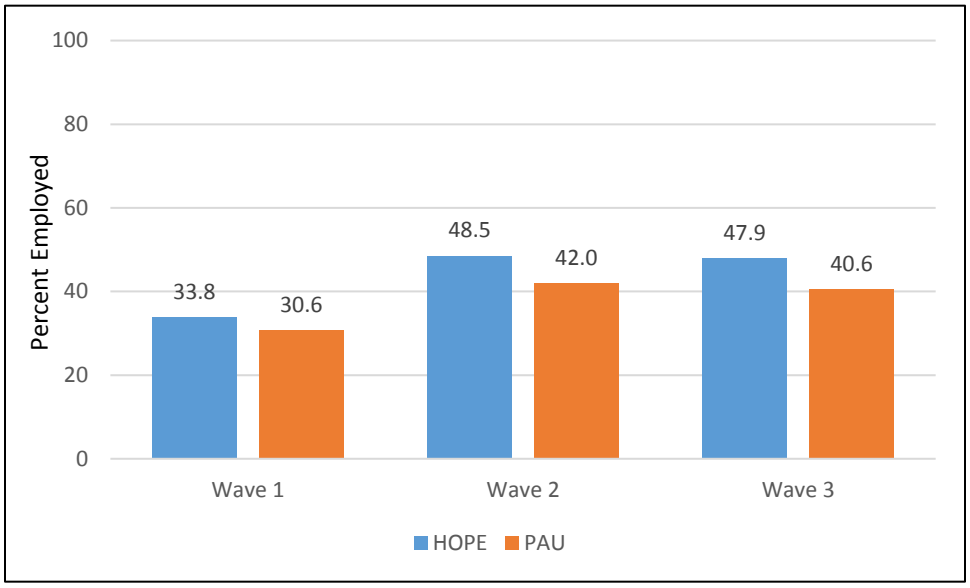
EMPLOYMENT

Individuals on HOPE and PAU reported similar rates of employment and similar wages across the three waves. HOPE probationers were more likely to have a job with formal pay at Wave 3 than PAU probationers.

Exhibit 3-16 presents information about the characteristics of probationers' current or most recent job. At the Wave 1 interview, about half of HOPE and PAU probationers reported an hourly wage between $10 and $20; about 40% in each group reported an hourly wage less than $10; and, about 10% reported an hourly wage more than $20. These rates were relatively stable across interview waves, and there were no significant differences between HOPE and PAU probationers. Across all interview waves, about 20% of HOPE probationers and about 25% of PAU probationers reported working more than 40 hours per week. Overall, there were no significant differences between HOPE and PAU probationers with respect to weekly hours worked. At the Wave 1 interview, PAU probationers were more likely than HOPE probationers to have a job with paid leave (32% versus 25%). At the Wave 2 and Wave 3 interview, the rates of employer-provided health insurance and paid leave among HOPE probationers were higher than but not significantly different from those among PAU probationers.

Across all interview waves, most probationers in both groups reported having a job with formal pay. ***At the Wave 3 interview, HOPE probationers were more likely than PAU probationers to report having a job with formal pay (77% versus 64%, respectively).***

Exhibit 3-15. Self-reported current employment status, by group and wave



Note: There were no differences between HOPE and PAU.

Exhibit 3-16. Self-reported characteristics of current or most recent job, by group and wave

Characteristics of current or most recent job	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
	HOPE	PAU	HOPE	PAU	HOPE	PAU
Hourly Wages						
Less than $10/hour	40.1	38.8	44.1	47.9	42.6	40
Between $10 and $20/hour	49.7	52.9	47.5	46.5	48.6	50.8
More than $20/hour	10.2	8.4	8.3	5.6	8.7	9.2
Weekly Hours						
More than 40 hours/week	21.2	25.8	19.4	21.5	21.1	26.7
40 hours/week (full-time)	38.2	32.6	44.7	35.4	41.6	34.2
Less than 40 hours/week (part-time)	40.6	41.6	35.9	43.1	37.3	39.2
Benefits						
Formal pay	71.2	69.8	71.8	66.0	76.6*	64.2
Health insurance	24.3	30.8	29.9	27.1	32.1	26.7
Paid leave	**24.8***	**32.3**	28.9	26.4	35.9	26.7

* HOPE and PAU differ at $p < 0.05$.

Sites also varied on employment indicators (Appendix G). At the Wave 1 interview, the employment rate ranged from 19% among probationers in Oregon, to 27% among probationers in Massachusetts, and 45% among probationers in Arkansas and Texas. There were no differences with respect to employment benefits or full-time/part-time status. There was more variation at Wave 2. For example, the current employment rate was 34% among Oregon probationers, 42% among Massachusetts probationers, 53% among Texas probationers, and 56% in Arkansas. Full-time status at their current or most recent job ranged from 26% of Oregon probationers to 47% of Arkansas probationers. Earnings and benefits also differed. Massachusetts had the highest proportion of probationers making more than $20 per hour and between $10 and $20 per hour at their current or most recent job (16% and 61%); it had the lowest proportion making less than $10 per hour (23%). Conversely, Texas had the lowest proportion of probationers making more than $20 per hour (0%) and between $10 and $20 per hour (38%); it had the highest proportion making less than $10 per hour (62%). With respect to benefits, 59% of Massachusetts probationers, 70% of Arkansas probationers, 74% of Texas probationers, and 77% of Oregon probationers reported that their current or most recent job provided formal pay. At the Wave 3 interview, current employment and earnings continued to vary by site. The current employment rate ranged from 28% among Oregon probationers to 60% among Texas probationers. Again, Massachusetts had the highest proportion of probationers earning more than $20 per hour (17%) and Texas had the lowest proportion (3%); Texas had the highest proportion of probationers earning less than $10 per hour, while Massachusetts had the lowest proportion (56% and 21%). At the Wave 3 interview, there were no differences among the sites with respect to benefits or full-time/part-time status.

Employment also varied by study group at the site level (***Exhibit 3-17*** and Appendix H), although there were no significant between-group differences in being currently employed at any wave. At Wave 1, more HOPE probationers in Arkansas than PAU probationers reported earning more than $20 per hour at their current or most recent job (7% versus 0%). Additionally, PAU probationers in Massachusetts were more likely than their HOPE counterparts to report that their current or most recent job provided health insurance (39% versus 25%). At Wave 2, the rate of employment among HOPE probationers was higher than but not significantly different from PAU probationers in three sites: Massachusetts, Oregon, and Texas (***Exhibit 3-17***). The only difference between study groups at Wave 2 was found in Oregon where more HOPE probationers than PAU probationers reported that they earned more than $20 per hour (10% versus 0%). At Wave 3, the rate of employment among HOPE probationers was higher than but not significantly different from PAU probationers in three sites: Arkansas, Massachusetts, and Texas. HOPE probationers in Arkansas were more likely than PAU probationers to report that their current or most recent job provided formal pay (75% versus 54%). The rate of formal pay among HOPE probationers was higher than but not significantly different from PAU probationers in Massachusetts and Oregon. PAU probationers in Texas were more likely than HOPE probationers to report that their current or most recent job provided formal pay (94% versus 77%). In Oregon, the rate of full-time employment among

EMPLOYMENT ACROSS SITES

Employment rates varied across site ranging at Wave 1 from 19% in Oregon and 27% in Massachusetts to 45% in Arkansas and Texas. Variation continued across the follow up waves with trends like the Wave 1 rates but higher in all sites.

HOPE probationers was higher than among PAU probationers (51% versus 24%). The rate of full-time employment among HOPE and PAU probationers was not significantly different in the other sites at Wave 3.

Exhibit 3-17. Self-reported current employment status (any employment), by site, group, and wave

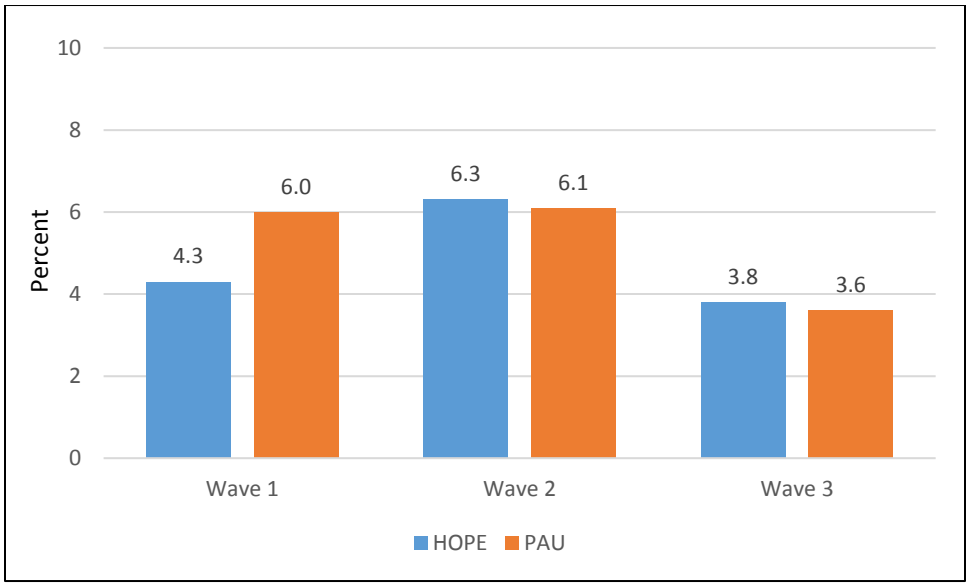
Site	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
	HOPE	PAU	HOPE	PAU	HOPE	PAU
Arkansas (Benton County)	41.2	50.0	55.3	56.9	58.0	44.6
Massachusetts (Essex County)	31.1	23.9	47.1	37.3	41.4	40.0
Oregon (Clackamas County)	17.1	20.6	34.3	32.8	26.2	31.4
Texas (Tarrant County)	47.7	41.9	57.7	46.2	63.5	52.0

There were no differences between HOPE and PAU.

Homelessness

Across interview waves, most HOPE and PAU probationers reported being housed in a place that they owned or rented, a group home, treatment facility, halfway house, a hotel, or with friends or family. ***Exhibit 3-18*** shows that at the Wave 1 interview, the rate of homelessness among HOPE probationers was somewhat lower than but not significantly different from PAU probationers (4% versus 6%). At the Wave 2 and Wave 3 interviews, HOPE and PAU probationers reported similar rates of homelessness.

Exhibit 3-18. Self-reported current homelessness, by group and wave



Note: There were no differences between HOPE and PAU.

At each interview wave, the homeless rate among probationers varied by site (Appendix G). The rate of homelessness was highest among Oregon probationers at the Wave 1 and Wave 2 interviews (10%

and 12%). At Wave 3, Oregon and Massachusetts probationers reported the highest rate of homelessness (6% and 7%, respectively). The rate of homelessness was lowest among Texas and Arkansas probationers at the Wave 1 interview (1%, and 2%), and at the Wave 3 interview (1%). At the Wave 2 interview, the homeless rate was lowest in Arkansas (2%). Although the homeless rate varied by site, it did not vary by study group within sites (Appendix H).

Current Mental Health

Probationers were asked to rate how they felt emotionally in the 30 days before their interview (***Exhibit 3-19***). The rating scale for item 1 and item 2 measured frequency as follows: 1=all of the time; 2=most of the time; 3=a good bit of the time; 4=some of the time; 5=a little of the time; 6=none of the time. The rating scale for item 3 and item 4 measured frequency as follows: 6=all of the time; 5=most of the time; 4=a good bit of the time; 3=some of the time; 2=a little of the time; 1=none of the time. The items were summed to arrive at a mental health symptom score where lower scores indicate lower symptom level. At the Wave 1 and Wave 2 interviews, HOPE and PAU probationers gave similar ratings to each item. At Wave 3, HOPE probationers felt emotionally better than PAU probationers with respect to the amount of time they had a lot of energy (3.0 versus 3.3, respectively) or felt down (2.8 versus 3.2, respectively). Overall, ***HOPE probationers reported a lower average mental health symptom level than PAU probationers at Wave 3 (11.2 versus 12.3).***

Exhibit 3-19. Self-reported level of mental health symptoms, by group and wave

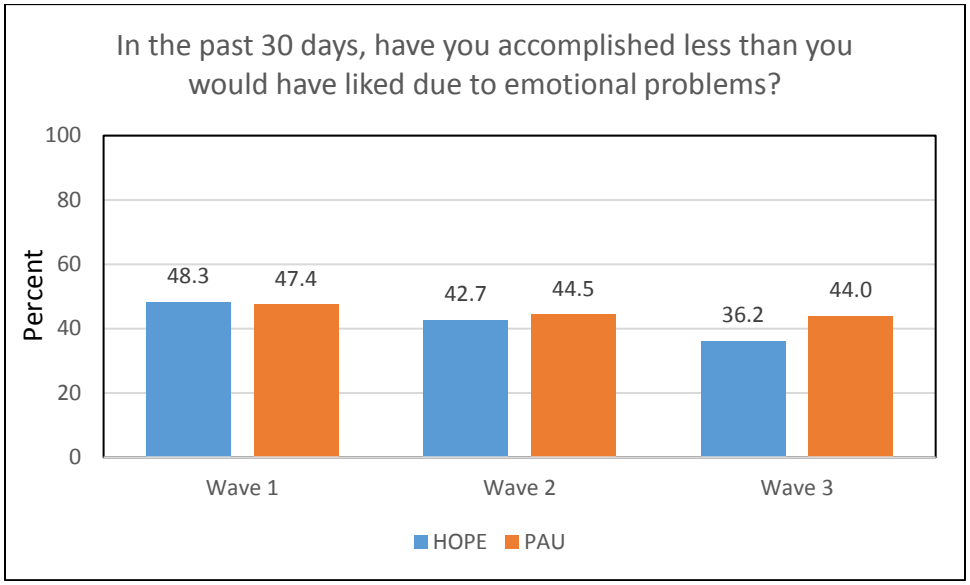
	Wave1 means		Wave 2 means		Wave 3 means	
In the past 30 days…	HOPE	PAU	HOPE	PAU	HOPE	PAU
1. How much time have you felt calm and peaceful?	3.3	3.4	3.3	3.3	3.0	3.1
2. How much time did you have a lot of energy?	3.2	3.1	3.1	3.3	**3.0***	**3.3**
3. How much time have you felt down?	3.1	3.1	3.1	3.1	**2.8****	**3.2**
4. How much time has your physical health or emotional problems interfered with social activities?	2.7	2.7	2.5	2.7	2.4	2.7
Mental health symptom score (range: 4 to 24)	12.2	12.2	12.0	12.3	**11.2***	**12.3**

**HOPE and PAU differ at p < 0.01; *HOPE and PAU differ at p < 0.05. Lower scores indicate lower symptom levels.
Note: The rating scale for items 1 and 2 measured frequency as follows: 1=all of the time; 2=most of the time; 3=a good bit of the time; 4=some of the time; 5=a little of the time; 6=none of the time. The rating scale for items 3 and 4 measured frequency as follows: 6=all of the time; 5=most of the time; 4=a good bit of the time; 3=some of the time; 2=a little of the time; 1=none of the time.

In addition to rating mental health symptoms, probationers were asked two questions related to the effect of any emotional problems they were experiencing. About 48% of both HOPE and PAU probationers said that they had accomplished less than they would have liked due to emotional problems in the 30 days before their Wave 1 interview (***Exhibit 3-20***). By Wave 3, the rate was somewhat lower among HOPE probationers but not significantly different from PAU probationers (36% versus 44%). ***Exhibit 3-21*** shows that when asked if they did not do work or activities as carefully as

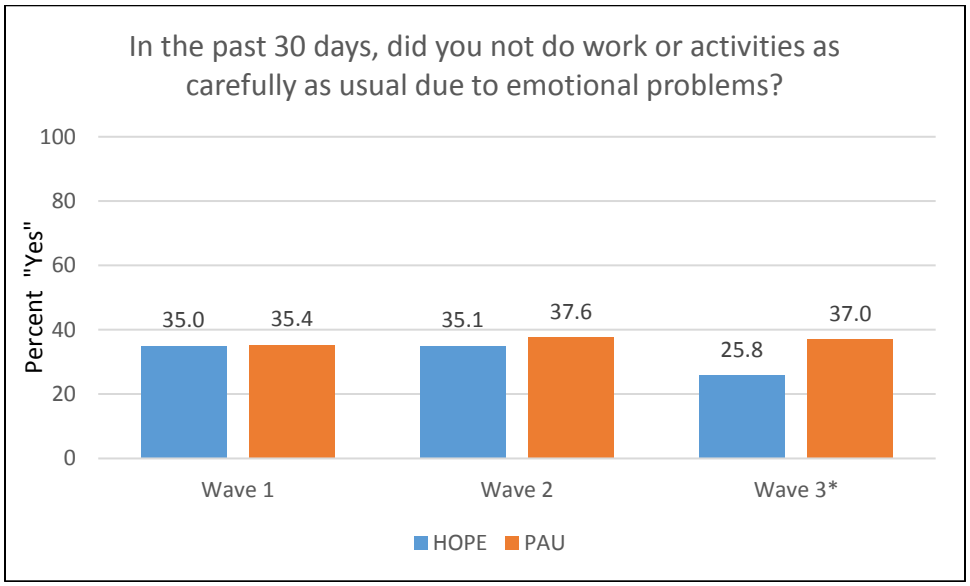
usual due to emotional problems in the 30 days before their Wave 1 interview, 35% of HOPE and PAU probationers said that they did not. By Wave 3, fewer HOPE probationers than PAU probationers said that they did not do work or activities as carefully as usual due to emotional problems (26% versus 37%).

Exhibit 3-20. Self-reported impact of emotional problems, by group and wave



Note: There were no differences between HOPE and PAU.

Exhibit 3-21. Self-reported impact of emotional problems on work, by group and wave



*HOPE and PAU differ at $p < 0.05$.

Mental health status varied across the sites (Appendix F). At all interview waves, probationers in Texas reported the lowest average mental health symptoms scores (11.3 at Wave 1, 10.9 at Wave 2, and 9.8 at Wave 3). At Wave 1 and Wave 2, probationers in Arkansas reported the highest average mental

health symptoms score (13.1 and 12.6). Probationers in Oregon reported the highest average mental health symptoms score at the Wave 3 interview (12.6).

At the Wave 1 interview, there were no differences between sites with respect to the effect of emotional problems in the past 30 days. However, differences were detected at the later interview waves. For example, 33% of Texas probationers, 42% of Massachusetts, 44% of Arkansas probationers, and 52% of Oregon probationers reported at Wave 2 that they had accomplished less than they would have liked due to emotional problems in the past 30 days. At Wave 3, 22% of Texas probationers, 26% of Arkansas probationers, 33% of Oregon probationers, and 40% of Massachusetts probationers reported that they had accomplished less due to emotional problems.

There were no differences in mental health symptom level between HOPE and PAU probationers at the Wave 1 and Wave 2 interviews (***Exhibit 3-22***). At Wave 3, HOPE probationers in Arkansas reported a lower level of symptoms than PAU probationers (10.4 versus 12.8). At the Wave 2 and Wave 3 interviews, HOPE probationers in Texas reported some lower mental health symptoms (at the item-level) than PAU probationers but their average symptoms score did not differ (Appendix H). With respect to the effect of past 30-day emotional problems, PAU probationers in Arkansas were more likely than their HOPE counterparts to report at the Wave 3 interview that they did not do work or activities as carefully as usual due to emotional problems (36% versus 18%). No other differences in the effects of past-30 day emotional problems were observed between study groups at the site level.

Exhibit 3-22. Self-reported level of mental health symptoms, by site, group, and wave

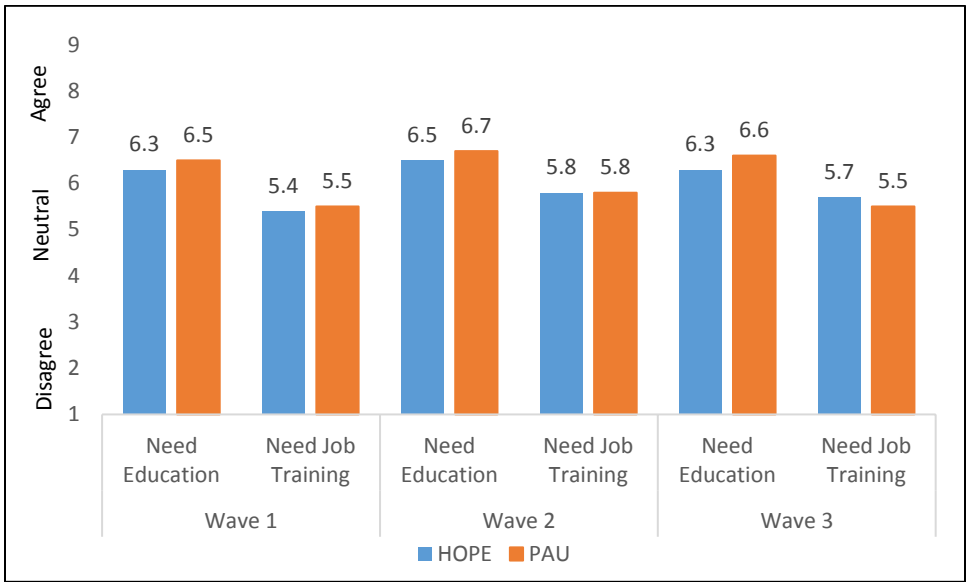
Site	Wave 1		Wave 2		Wave 3	
	HOPE mean	PAU mean	HOPE mean	PAU mean	HOPE mean	PAU mean
Arkansas (Benton County)	13.3	12.8	12.4	12.8	**10.4****	**12.8**
Massachusetts (Essex County)	12.2	12.4	12.0	12.2	12.8	12.1
Oregon (Clackamas County)	12.4	12.4	12.7	12.5	12.7	12.4
Texas (Tarrant County)	11.1	11.5	10.2	11.8	9.2	11.1

** HOPE and PAU differ at $p < 0.01$. Lower scores indicate lower symptom levels.

Current Service Needs and Receipt

At each interview wave, participants were asked how much they agreed that they needed a variety of services including drug or alcohol treatment, mental health treatment or mental health care, education, and job training. Level of agreement for need of each service was captured on a 9-point scale where 1=strongly disagree and 9=strongly agree. The HOPE and PAU probationers provided similar responses with respect to need. On average, HOPE and PAU probationers agreed that they needed education and neither agreed or disagreed that they needed job training across the waves (***Exhibit 3-23***).

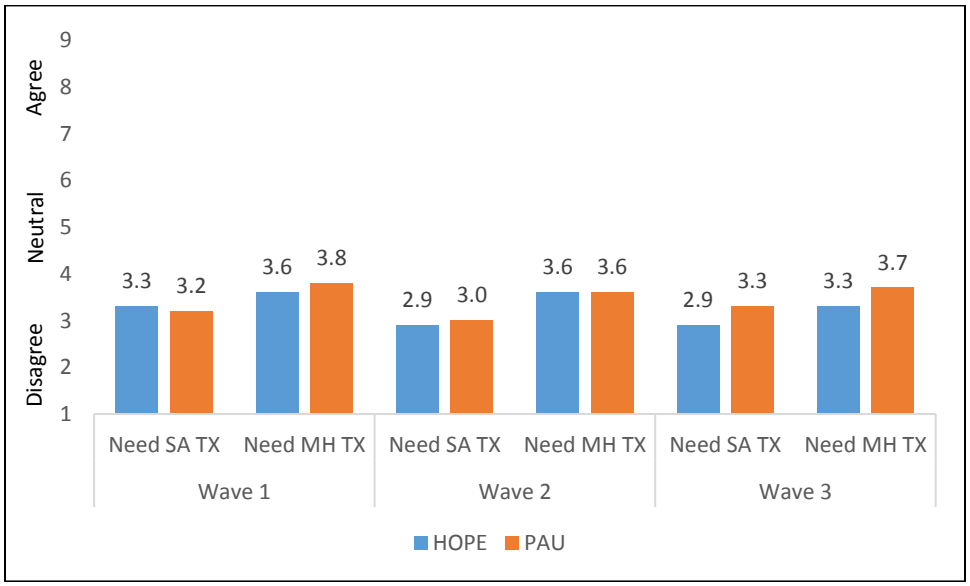
Exhibit 3-23. Self-reported need for education and employment services, by group and wave



Note: There were no differences in employment status between HOPE and PAU.

Exhibit 3-24 shows that, on average, ***probationers in both groups disagreed that they needed drug or alcohol treatment***. Across interview waves, there were no significant differences between groups with respect to their need for drug or alcohol treatment. Similar findings were found with respect to reported need for mental health treatment, with both groups disagreeing that they needed mental health treatment.

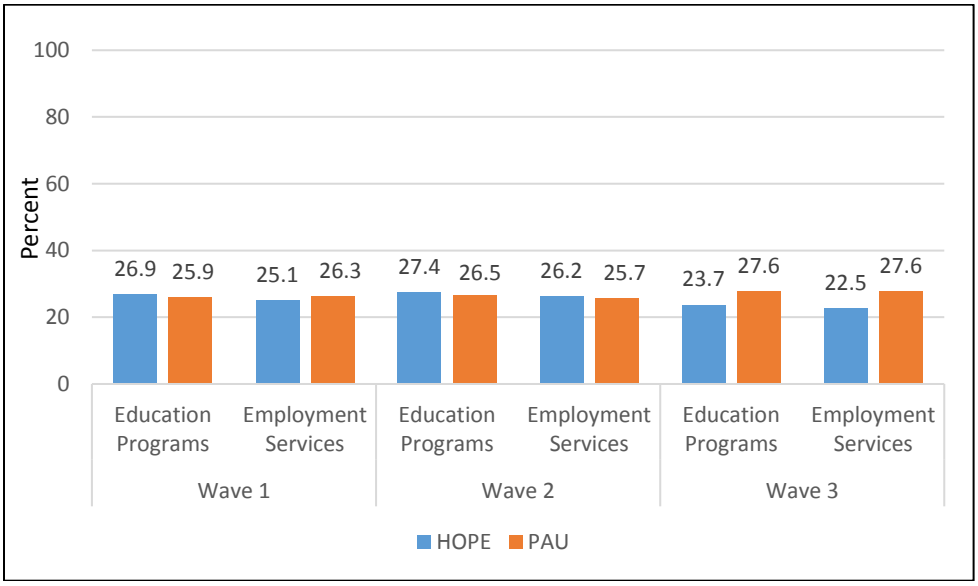
Exhibit 3-24. Self-reported need for behavioral health care, by group and wave



Note: There were no differences between HOPE and PAU.

At each interview, participants were asked about their participation in education programs such as GED or adult basic education classes, and receipt of employment services such as job training, employment services or assistance with finding a job over the past 6 months. At Wave 1 and 2 interviews, about one-quarter of HOPE and PAU probationers reported participating in education programs and receiving employment services (***Exhibit 3-25***). Compared to PAU probationers, HOPE probationers reported a lower rate of participation in education programs (28% versus 24%) and receipt of employment services (28% versus 22%) at Wave 3, but the differences were not significant.

Exhibit 3-25. Self-reported past 6-month education and employment service receipt, by group and wave



Note: There were no differences between HOPE and PAU.

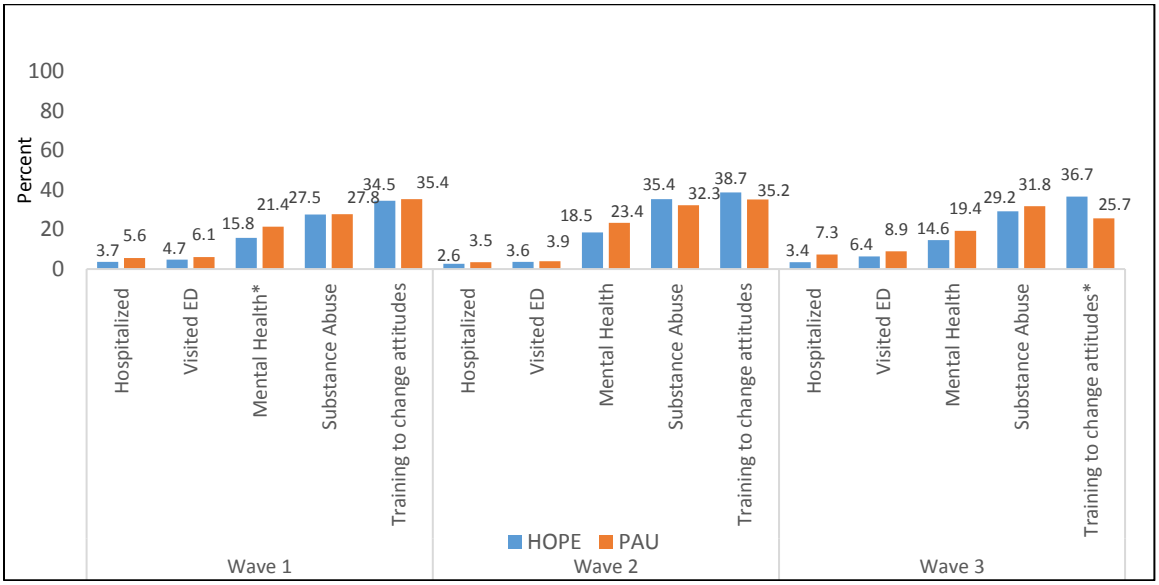
At the Wave 1 and Wave 2 interviews, there were no differences among sites with respect to probationers' reported need of education and job training (Appendix G). At Wave 3, reported need of education among probationers varied by site. Probationers in Oregon, Texas, and Arkansas reported a higher level of agreement with the need for education than probationers in Massachusetts. Across sites, there was no difference in the reported need for job training at the Wave 3 interview.

Across interview waves, participation in education programs and receipt of employment services in the past 6 months varied by site (Appendix G). At the Wave 1 interview, participation in education programs was highest among probationers in Massachusetts and lowest among probationers in Arkansas (32% and 16%). Arkansas probationers continued to report the lowest participation rates at the Wave 2 and Wave 3 interviews (18%). Probationers in Oregon reported the highest rate of participation at the Wave 2 and Wave 3 interviews (36% and 32%). With respect to receipt of employment services, the rate of receipt of these services was highest among probationers in Texas and lowest among probationers in Arkansas at the Wave 1 interview (33% and 20%). At the Wave 2 interview, receipt of employment services was highest among probationers in Oregon and Texas (33% and 38%) and lowest among probationers in Arkansas and Massachusetts (19%). The receipt of employment services reported at the Wave 3 interview followed the same pattern as the Wave 2

interview. Across all interview waves, there were no differences between HOPE and PAU probationers at the site level with respect to their reported need for education or job training (Appendix H).

Participants were also asked about their past 6-month receipt of services including drug or alcohol treatment, treatment in an emergency department due to an alcohol or drug problem, treatment in a hospital due to an alcohol or drug problem, mental health treatment or health care for emotional problems, and training on how to change attitudes related to criminal behavior. Despite disagreeing with the need for behavioral health services, interview participants reported receiving these services. ***Exhibit 3-26*** shows that, ***at Wave 1, more than one-quarter of probationers in each group reported receiving alcohol or drug treatment in the past 6 months*** (27% of HOPE and 28% of PAU). Additionally, ***21% of PAU probationers reported receipt of mental health treatment in the past 6 months, a higher rate than reported by HOPE probationers (16%).*** Alcohol or drug problems resulted in hospitalization in the past 6 months for 4% of HOPE probationers and 6% of PAU probationers. Similarly, 5% of HOPE probationers and 6% of PAU probationers had visited an emergency department (ED) in the past 6 months due to an alcohol or drug problem. ***More than one-third of probationers in each group reported receiving training on how to change their attitudes related to their criminal behavior (35%).*** At the Wave 2 interview, past 6-month hospitalizations and ED visits remained low for both groups. Receipt of mental health care in the 6 months before the Wave 2 interview was higher for PAU probationers than for HOPE probationers, but the difference was not significant (23% versus 18%, respectively). Past 6-month receipt of alcohol or drug treatment was higher for HOPE probationers than for PAU probationers, but the difference was not significant (35% versus 32%). At the Wave 3 interview, the rate of past 6-month service receipt among PAU probationers was higher but not significantly different from HOPE probationers, except for receipt of training on how to change attitudes related to criminal behavior where receipt was higher among HOPE probationers than PAU probationers (37% versus 26%).

Exhibit 3-26. Self-reported past 6-month behavioral health service receipt, by group and wave



* HOPE and PAU differ at $p < 0.05$.

At each interview, participants who received behavioral health treatment in the past 6 months were asked what type of treatment they received (***Exhibit 3-27***). At Wave 1, HOPE and PAU probationers most frequently reported that they participated in AA/NA (60% and 61%), followed by drug education classes (48% and 50%) and group counseling (45%). About one-quarter of probationers in each group reported that they had received individual counseling in the past 6 months. Residential substance abuse treatment and detox were less frequent than other substance abuse services for both groups. The rate of inpatient mental health treatment among HOPE probationers was higher than that among PAU probationers, but the difference was not significant (36% and 26%). At Wave 2, the rates of participation in AA/NA and drug education class, as well as receipt of individual counseling and residential substance abuse treatment among HOPE probationers were higher than but not significantly different from PAU probationers. The rate of detox among HOPE probationers was nearly twice that among PAU probationers (12% and 7%). The rate of inpatient mental health treatment among HOPE probationers was higher than that among PAU probationers, but the difference was not significant (32% and 28%). At Wave 3, the rate of participation in AA/NA was higher among HOPE probationers than PAU probationers (76% versus 56%), as was the receipt of residential substance abuse treatment (38% versus 20%).

Exhibit 3-27. Self-reported past 6-month behavioral health treatment among those receiving treatment, by group and wave

	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
Type of Treatment	HOPE	PAU	HOPE	PAU	HOPE	PAU
AA/NA	59.7	61.2	75.7	66.2	**75.6***	**55.7**
Group counseling	45.5	44.8	54.2	55.4	52.6	41.0
Individual counseling	23.9	25.4	41.1	37.8	35.9	34.4
Detox	16.4	19.4	12.1	6.8	9.0	8.2
Drug education classes	47.8	50.0	48.6	40.5	41.0	52.5
Residential substance abuse Tx	17.2	19.4	36.4	23.0	**38.5***	**19.7**
Inpatient mental health Tx	36.4	26.2	32.1	27.8	28.2	24.3
N	134	134	107	74	78	61

* HOPE and PAU differ at $p < 0.05$.

The reported need for drug or alcohol treatment and mental health treatment varied by site at all interview waves (Appendix G). ***Although probationers generally disagreed that they needed mental health treatment, probationers in Massachusetts and Oregon reported higher levels of agreement with the need for treatment than probationers in the other two sites.*** This was true across all interview waves. ***With respect to the need for drug or alcohol treatment, probationers in Oregon generally reported higher levels of agreement with the need for treatment at all interview waves.***

Receipt of behavioral health services in the past 6 months also varied by site across all interview waves (Appendix G). ***Receipt of drug or alcohol treatment in the past 6 months ranged from a low of 13% of probationers in Arkansas to a high of 35% of probationers in Massachusetts at the Wave 1 interview.*** At the Wave 2 interview, Massachusetts probationers reported the lowest rate of treatment receipt, and probationers in Oregon reported the highest rate of receipt (26% and 50%). Probationers in

Oregon continued to report the highest rate of treatment receipt at the Wave 3 interview, while Arkansas probationers reported the lowest rate (41% and 15%). At the Wave 1 interview, the rate of mental health treatment receipt ranged from a high of 30% of probationers in Massachusetts to a low of 8% of probationers in Texas. At the Wave 2 interview, probationers in Arkansas and Texas reported similar low rates of treatment receipt (12%). Probationers in Massachusetts reported the highest rate of treatment receipt at the Wave 2 interview (30%). At the Wave 3 interview, the rate of treatment receipt ranged from 26% of probationers in Massachusetts to 9% of probationers in Arkansas.

Among probationers receiving behavioral health services in the past six months, the type of treatment they received varied by site (Appendix G). For example, receipt of group counseling was highest among probationers in Oregon and lowest among probationers in Massachusetts and Arkansas at the Wave 1 interview (60%, 37%, and 36%). Receipt of residential substance abuse treatment ranged from 4% of probationers in Texas to 32% of probationers in Arkansas. At the Wave 2 interview, participation in AA/NA was the most commonly reported type of treatment in Arkansas (71%), Massachusetts (59%), and Oregon (89%), with drug education/classes most common in Texas (66%). varied but was common in all sites with highest participation among probationers in Oregon and lowest among probationers in Texas (89% and 56%). Participation in drug education classes varied but was also quite common with 66% of probationers in Texas, 52% of probations in Arkansas, 43% of probationers in Oregon, and 27% of probationers in Massachusetts reporting participation. Receipt of group counseling ranged from 34% among probationers in Massachusetts to 67% among probationers in Oregon. There was a wide range in the rate of residential substance abuse treatment receipt among those who reported treatment, from 16% of probationers in Massachusetts to 48% of probationers in Arkansas. At Wave 3, there was less variation among sites. Rates of receipt varied by site only for group counseling and residential substance abuse treatment.

HOPE and PAU probationers persisted in disagreeing that they needed drug or alcohol treatment or mental health treatment across all interview waves (Appendix H). The only significant difference between HOPE and PAU probationers was detected in Texas at the Wave 1 interview where HOPE probationers reported a higher level of disagreement than PAU with the need for mental health treatment (2.4 versus 3.3).

With respect to service receipt in the past 6 months, there were no differences between HOPE and PAU probationers at the site level at the Wave 1 interview (***Exhibit 3-28***). Service receipt among HOPE and PAU probationers was similar at the Wave 2 and Wave 3 interviews with a few exceptions. At the Wave 2 interview, HOPE probationers in Arkansas were more likely than PAU probationers to report that they received drug or alcohol treatment in the past 6 months (38% versus 12%). HOPE probationers in Arkansas were also more likely than PAU probationers to report that they had received training on how to change attitudes related to criminal behavior (32% versus 16%). In Oregon, PAU probationers were more likely than their HOPE counterparts to report that they had received drug or alcohol treatment in the past 6 months (61% versus 41%). At the Wave 3 interview, PAU probationers in Texas were more likely than HOPE probationers to report that they received job training, employment services, or help finding a job (52.0% versus 20.6%, respectively).

There were also a few differences with respect to the type of behavioral health service received among those receiving treatment between HOPE and PAU probationers at the site level (Appendix H). At the Wave 2 interview, HOPE probationers in Arkansas were more likely than PAU probationers to report

that they had received detox services (14% versus 0%). At the Wave 3 interview, HOPE probationers in Arkansas were more likely than PAU probationers to report that they received residential substance abuse treatment (69% versus 12%). In Oregon, PAU probationers were more likely than HOPE probationers to report that they had participated in drug education classes (70% versus 39%).

Exhibit 3-28. Self-reported past 6-month service receipt, by site, group, and wave

	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
Site	HOPE	PAU	HOPE	PAU	HOPE	PAU
Education programs						
Arkansas (Benton County)	17.0	13.4	14.9	23.5	17.3	19.6
Massachusetts (Essex County)	32.5	32.0	28.7	26.5	27.6	31.7
Oregon (Clackamas County)	27.0	17.8	41.4	29.3	32.3	31.4
Texas (Tarrant County)	27.3	33.0	28.8	26.3	19.4	28.0
Employment Services						
Arkansas (Benton County)	19.8	19.4	19.1	17.6	19.8	21.4
Massachusetts (Essex County)	25.2	22.1	21.8	16.9	15.5	21.7
Oregon (Clackamas County)	25.2	27.7	41.4	34.5	33.8	29.4
Texas (Tarrant County)	29.7	36.5	25.5	42.1	20.6**	52.0
Drug or alcohol treatment						
Arkansas (Benton County)	13.9	12.1	**38.3****	**11.8**	16.0	14.3
Massachusetts (Essex County)	35.6	34.3	25.3	26.5	29.3	40.0
Oregon (Clackamas County)	30.4	32.3	**40.6***	**61.4**	43.1	39.2
Texas (Tarrant County)	25.2	21.0	40.4	28.9	31.7	36.0
Mental health treatment or mental health care						
Arkansas (Benton County)	7.9	18.5	10.6	15.7	6.2	12.5
Massachusetts (Essex County)	27.6	32.6	31.0	28.6	25.9	27.1
Oregon (Clackamas County)	16.1	16.2	17.4	31.0	16.9	15.7
Texas (Tarrant County)	5.4	10.5	13.5	10.5	12.7	24.0
Training on how to change attitudes related to criminal behavior						
Arkansas (Benton County)	12.9	24.2	**31.9***	**15.7**	25.9	16.1
Massachusetts (Essex County)	46.0	40.3	31.0	35.7	22.4	25.0
Oregon (Clackamas County)	30.4	36.4	46.4	38.6	46.2	33.3
Texas (Tarrant County)	41.4	32.7	53.8	55.3	54.0	33.3

** HOPE and PAU differ at $p < 0.01$; * HOPE and PAU differ at $p < 0.05$.

Criminal History

Exhibit 3-29 shows self-reported criminal justice system involvement among HOPE and PAU probationers. Across all interview waves, there were no significant differences between HOPE and PAU probationers in their reported lifetime arrests or convictions. At the Wave 1 interview, HOPE probationers reported an average of 9.6 arrests, compared to 10.4 arrests reported by PAU

probationers. At the Wave 2 interview, an average of 12.2 arrests were reported by HOPE probationers, compared to 10.4 arrests reported by PAU probationers. At Wave 3, HOPE probationers reported an average of 12.6 arrests in their lifetimes, while PAU probationers reported 11.6 arrests. Both groups consistently reported having about 5 lifetime convictions.

Exhibit 3-29 also shows that HOPE probationers and PAU probationers reported a similar average number of lifetime incarcerations at the Wave 1 interview (5.9 and 5.8). ***The Wave 2 average number of incarcerations reported by HOPE probationers was significantly higher than that reported by PAU probationers (7.9 versus 6.3)—perhaps reflecting the use of jail as a sanction for HOPE. The average number of incarcerations reported by HOPE probationers in Wave 3 was also significantly higher than that reported by PAU probationers (8.3 versus 6.7).*** When asked how many days they had been incarcerated in the past 6 months, HOPE probationers and PAU probationers reported similar experiences at the Wave 1 interview (45.1 days and 42.2 days, respectively). ***At the Wave 2 interview, HOPE probationers reported a higher average number of days spent incarcerated than did PAU probationers (35.2 days versus 23.0 days). HOPE probationers also reported a higher average number of days spent incarcerated at the Wave 3 interview than did PAU probationers (40.7 days versus 29.5 days).***

INCARCERATION HISTORIES

HOPE and PAU probationers reported similar numbers of prior incarcerations at baseline. During follow up interviews, HOPE probationers reported more lifetime incarcerations and more days incarcerated in the prior 6 months than PAU probationers.

Exhibit 3-29. Self-reported criminal history, by group and wave

	Wave1 means		Wave 2 means		Wave 3 means	
Criminal History	HOPE	PAU	HOPE	PAU	HOPE	PAU
Lifetime arrests	**9.6**	**10.4**	12.2	10.4	12.6	11.6
Lifetime convictions	**4.7**	**4.9**	**5.3**	**5.0**	4.7	5.2
Lifetime incarcerations	5.9	5.8	**7.9***	**6.3**	**8.3***	**6.7**
Days incarcerated in the past 6 months	45.1	42.2	**35.2***	**23.0**	**40.7***	**29.5**

* HOPE and PAU differ at $p < 0.05$.

Criminal histories of probationers varied by site (Appendix G). At Wave 1, probationers in Oregon reported the highest average number of lifetime arrests (13.5) and highest average number of lifetime incarcerations (9.4). Probationers in Massachusetts had the highest number of lifetime convictions (5.8), but the lowest number of lifetime incarcerations (4.3). Texas probationers had the lowest lifetime arrests and convictions (6.7 and 2.9). At Wave 2, probationers in Oregon reported the highest lifetime arrests, convictions, and incarcerations (17.4, 6.4, and 11.9). Probationers in Texas had the lowest lifetime arrests and convictions (8.4 and 3.3). The lowest lifetime incarcerations were reported by Massachusetts probationers (5.2). At Wave 3, probationers in Oregon had the highest lifetime arrests and incarcerations (17.2 and 12.5). Massachusetts probationers had the highest lifetime convictions

(6.2), but lowest lifetime incarcerations (5.1). Probationers in Texas had the lowest lifetime arrests and convictions (9.5 and 2.8).

Exhibit 3-30 shows criminal history among probationers by site, group, and interview wave. At Wave 1, HOPE and PAU probationers within sites shared similar criminal history characteristics. One exception was found in Texas, where HOPE probationers had more lifetime convictions on average than PAU probationers (3.3 versus 2.4). Another exception was found in Arkansas, where HOPE probationers reported more average days incarcerated in the past 6 months than PAU probationers (24.2 versus 10.4, respectively).

Exhibit 3-30. Self-reported criminal history, by site, group, and wave

	Wave 1 means		Wave 2 means		Wave 3 means	
Site	**HOPE**	**PAU**	**HOPE**	**PAU**	**HOPE**	**PAU**
Lifetime Arrests						
Arkansas (Benton County)	7.6	7.6	8.9	8.3	9.7	9.7
Massachusetts (Essex County)	10.6	10.9	12.0	10.5	14.3	10.6
Oregon (Clackamas County)	12.9	14.1	**19.6***	**15.1**	18.6	15.5
Texas (Tarrant County)	6.7	6.9	**9.9***	**6.4**	9.1	10.3
Lifetime Convictions						
Arkansas (Benton County)	3.8	3.8	4.5	4.3	4.0	4.7
Massachusetts (Essex County)	5.4	6.3	6.0	5.6	7.1	5.2
Oregon (Clackamas County)	5.6	5.3	6.8	5.9	5.5	6.2
Texas (Tarrant County)	**3.3***	**2.4**	3.6	3.0	2.4	4.2
Lifetime Incarcerations						
Arkansas (Benton County)	5.3	5.1	6.5	5.2	6.4	7.2
Massachusetts (Essex County)	4.4	4.1	5.4	4.9	5.8	4.4
Oregon (Clackamas County)	9.4	9.5	13.2	10.4	**14.3***	**10.4**
Texas (Tarrant County)	4.5	4.5	**8.0***	**4.6**	**7.1****	**3.0**
Days Incarcerated in Past Six Months						
Arkansas (Benton County)	**24.2****	**10.4**	**42.1****	**10.8**	40.5	25.4
Massachusetts (Essex County)	67.8	62.6	19.6	17.5	30.0	21.3
Oregon (Clackamas County)	56.7	49.2	48.5	44.6	53.0	39.0
Texas (Tarrant County)	19.1	17.9	32.1	19.4	38.7	39.0

** HOPE and PAU differ at $p < 0.01$; * HOPE and PAU differ at $p < 0.05$.

At the Wave 2 interview, HOPE probationers in Oregon and Texas had more lifetime arrests on average than their PAU counterparts (in Oregon, 19.6 versus 15.1; in Texas, 9.9 versus 6.4). In addition, HOPE probationers in Texas had more average lifetime incarcerations than PAU probationers (8.0 versus 4.6). In Arkansas, HOPE probationers reported more average days incarcerated in the past six months than PAU probationers (42.1 versus 10.8). HOPE probationers in the other three study sites also reported more average days incarcerated in the past 6 months than PAU probationers, but the

differences were not significant. At the Wave 3 interview, HOPE probationers in Oregon and Texas had more average lifetime incarcerations than PAU probationers in these sites (in Oregon, 14.3 versus 10.4; in Texas, 7.1 versus 3.0). In three of the four study sites (Arkansas, Massachusetts, and Oregon), HOPE probationers reported more average days incarcerated in the past 6 months than PAU probationers, but the differences in these sites were not significant.

Family and Peers

Exhibit 3-31 presents information about criminal justice involvement and substance use among families of HOPE and PAU probationers. ***Across all interview waves, most of HOPE and PAU probationers reported that they had a family member who had been convicted, had been incarcerated, or had alcohol or drug problems.*** More than 20% of probationers in each group at Wave 1 and about 25% of probationers in each group at Wave 2 and Wave 3 reported that they had a family member who was currently incarcerated. Overall, there were no significant differences between HOPE and PAU probationers with respect to family criminal justice involvment or substance use.

Exhibit 3-31. Self-reported criminal justice involvement and substance use by family, by group and wave

Family Criminal Justice Involvement and Substance Use	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
	HOPE	PAU	HOPE	PAU	HOPE	PAU
Any family member ever convicted	53.6	57.4	55.0	63.5	61.2	56.3
Any family member ever incarcerated+	55.5	59.2	55.6	62.7	59.5	56.9
Any family member currently incarcerated+	22.2	21.5	24.8	24.4	26.1	24.1
Any family member with alcohol or drug problems	59.9	61.7	62.7	68.2	62.4	66.9

There were no differences between HOPE and PAU. + Jail, prison, or juvenile detention facility.

Exhibit 3-32 presents information about criminal justice involvement and substance use among peers of HOPE and PAU probationers. At Wave 1, 16% of HOPE and 20% of PAU probationers reported that most or all of their close friends have problems with alcohol or drugs. Similarly, 13% of HOPE probationers and 14% of PAU probationers reported that most or all of their close friends are frequently drunk or high. At Wave 1, about 20% of both groups reported that most or all of their close friends had been convicted and more than 20% in each group reported that most or all of their close friends had been incarcerated, with about 6% in each group reporting that most or all of their close friends were currently incarcerated. ***By Wave 3, fewer HOPE probationers than PAU probationers reported that most or all of their friends are frequently drunk or high (10% versus 20%). Additionally, fewer HOPE probationers reported that most or all close friends had been incarcerated (20% versus 31%).***

Exhibit 3-32. Self-reported criminal justice involvment and substance use among peers, by group and wave

Peer Criminal Justice Involvement and Substance Use	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
	HOPE	PAU	HOPE	PAU	HOPE	PAU
Most or all close friends are frequently drunk or high	12.8	14.3	14.0	14.5	**10.3****	**19.6**
Most or all close friends have problems with alcohol or drugs	16.4	19.6	23.3	20.6	17.9	24.3
Most or all close friends are currently incarcerated	6.5	6.0	5.7	4.4	5.7	5.3
Most or all close friends have been incarcerated	23.0	21.6	23.6	30.8	**20.2***	**30.7**
Most or all close friends have been convicted	18.9	20.6	23.1	24.2	20	27.4

** HOPE and PAU differ at p < 0.01; * HOPE and PAU differ at p < 0.05.

Sites varied with respect to family criminal justice involvement, but not family substance use (Appendix G.) For example, ***the majority of probationers in each site reported having a family member who had been incarcerated, but the proportion of probationers having a family member with incarceration history ranged from 51% in Massachusetts to 70% in Texas at Wave 1.*** Texas also had the highest proportion of probationers with a family member currently incarcerated (33.3%), while Oregon had the lowest (16%). At Wave 2, Oregon and Texas had the highest proportion of probationers having a family member with a conviction history (66%). Texas also had the highest proportion of probationers having a family member with incarceration history and one who was currently incarcerated (72% and 43%). Massachusetts had the lowest proportion of probationers having a family member with conviction history and incarceration history (48% and 46%). Oregon had the lowest proportion of probationers having a family member currently incarcerated (17%). At Wave 3, there were no site level differences in family criminal justice involvement.

Sites also varied with respect to peer criminal justice involvement and peer substance use (Appendix G). ***At all interview waves, Oregon probationers ranked highest in the proportion reporting that most or all close friends had been incarcerated***: 29% at Wave 1, 35% at Wave 2, and 36% at Wave 3. At all interview waves, the lowest proportion of probationers reporting that most or all their close friends had been incarcerated was in Arkansas (18% at Wave 1, 20% at Wave 2, 18% at Wave 3) and Texas (18% at Wave 1, 20% at Wave 2, 14% at Wave 3). At the Wave 2 and Wave 3 interviews, Texas probationers ranked lowest in the proportion reporting that most or all close friends had been convicted and in the proportion reporting that most or all close friends are frequently drunk or high. At Wave 3, Texas probationers also ranked the lowest in the proportion reporting that most or all their close friends have problems with drugs or alcohol (11.5%). At Wave 2, Oregon probationers ranked highest in the proportion reporting that most or all close friends had been convicted (32.5%) and in the proportion reporting that most or all close friends are frequently high (21%). At Wave 3, Oregon probationers also ranked highest in the proportion reporting that all or most of their close friends had been convicted (34%) and in the proportion reporting that most or all their close friends had problems with drugs or alcohol (32%). At the Wave 3 interview, the highest proportion of probationers reporting that most or all their close friends are frequently drunk or high was in Massachusetts (20%).

At the Wave 1 interview, there were no difference between HOPE and PAU probationers at the site level with respect to criminal justice involvement or substance use among family and friends (Appendix

H). At the Wave 2 interview, PAU probationers in Texas were more likely than HOPE probationers to report having a family member who had been convicted (82% versus 53%), and to report having a family member who had been incarcerated (85% versus 63%). In Oregon, HOPE probationers were more likely than PAU probationers to report that they had a family member who was currently incarcerated (24% versus 8%), and that most or all their close friends were currently incarcerated (9% versus 0%). PAU probationers in Massachusetts were more likely than HOPE probationers to report having a family member that had problems with drugs or alcohol (71% versus 52%). At Wave 3, PAU probationers in Arkansas were more likely than their HOPE counterparts to report that most or all their close friends have been incarcerated (27% versus 11%).

3.4. Summary

A total of 1,580 HOPE-eligible probationers were randomly assigned to either HOPE or PAU in the four DFE sites. Of these, 76 individuals were subsequently determined to by study ineligible, leaving a final sample of 1,504 (743 assigned to HOPE, 761 assigned to PAU). Most were male (81%), white (69%), and high risk (55%). On average, they were 31 years at study enrollment, with 7 prior arrests and 3.5 prior convictions. Most were on probation for either a drug (31%) or property (30%) offense.

Subject characteristics varied across sites, but generally not between groups within sites. For example, study participants were younger at first arrest in Texas than Arkansas (19 versus 27 years) and the number of prior convictions ranged from 1.7 in Arkansas to 5.8 in Massachusetts. While the number of prior arrests was about 7 overall, average numbers of prior arrests was 13.0 in Massachusetts, while ranging between 4.4 and 6 in the other three sites.

Interviews conducted at study intake and at 6 and 12 months post-intake showed few differences between HOPE and PAU on a variety of measures. For example, probationers on HOPE and PAU reported similar rates of employment and similar wages. About 40% of both groups reported accomplishing less than they would have liked because of emotional problems across all three waves.

Both groups were similar in their assessment of need for specific services or treatment. Members of both groups agreed that they needed more education but not necessarily more job training. Groups reported receiving similar levels of education and employment services.

Neither group believed that they needed substance abuse or mental health treatment. However, HOPE probationers were more likely to report having received residential substance abuse treatment at the 6- and 12-month interviews. At the 12-month interview, HOPE probationers were less likely than PAU probationers to report that most or all their close friends are frequently drunk or high.

In summary, HOPE and PAU groups were similar across sites and within sites. However, there was heterogeneity across sites with respect to some key metrics, including extent of criminal history and age.

4. Findings from the Process Evaluation: Fidelity and Implementation

The process evaluation addressed the following research questions:

1. What was the structural context for the implementation of HOPE in the four sites?
2. Was HOPE implemented with fidelity in the four sites?
3. What lessons were learned for implementation success, replicability, and sustainability?
4. How do intensive drug treatment services offered with the HOPE programs compare with the principles of effective offender intervention?
5. What were the communication pathways among HOPE stakeholders and did these vary from site to site?
6. How did HOPE probationers view their supervision experiences?

4.1. Structural Context for HOPE Programs

HOPE programs require cooperation between multiple key stakeholders, including the HOPE judge, probation officers and management, the district attorney or local prosecutor, the public defender or local defense bar, local law enforcement officers or units for warrant service, and local jail management (usually the Sheriff) for sanctioning. The importance of each of these individuals or entities varies across level of government (state and local) and branch of government (executive and judicial).

The four sites selected for the DFE varied along these dimensions and these variations along three key administrative dimensions shaped the implementation of HOPE. The following is adapted from Zajac, Lattimore, Dawes, and Winger (2015).

The first administrative dimension concerns how probation is organized – in other words, who controls probation/community corrections. In two of the sites (Saline County, Arkansas, and Tarrant County, Texas), probation is an independent executive agency administered at either the state level or at a hybrid of state-county levels. In a third site (Essex County, Massachusetts), probation is subsumed under a larger state court administrative office and thus is directly part of the judiciary (much as in Hawaii HOPE). At the fourth site (Clackamas County, Oregon), probation is administered through the county Sheriff department although funding is provided by the state. ***This organizational dimension affects how the probation office is affiliated with other key HOPE stakeholders—in particular, whether there is a formal administrative linkage between probation and the court or whether the implementation and operation of the HOPE program must rely on robust informal relationships between these key stakeholders to the HOPE program.*** Other jurisdictions considering adopting HOPE should consider the strength and formality of these arrangements as they devise their implementation strategy.

The second related dimension concerns the degree of control that the HOPE judge can exercise over the HOPE probationer officers. HOPE is by design a judge-driven model, so the ability of the judge to direct the work of the HOPE probation officers should in principle be an important implementation variable. In Saline County, Essex County, and Tarrant County, the HOPE judges seemed to exercise a high degree of formal or informal direction over the work of the HOPE probation officers with respect to tasks such as how swiftly violations were responded to, strict compliance with drug testing regimens, keeping probationers apprised of the requirement of HOPE, and other HOPE-related supervision practices. In Clackamas County, the judge seemed to have relatively less direct control over the HOPE

probation officers. Thus, the core issue here is the judge's ability to ensure that a bright line is maintained between HOPE supervision and supervision as usual. This was clearly an issue for the implementation of our evaluation (e.g. treatment group versus control group conditions), but it is also critical to the management of the program, particularly regarding the ability of the judge to ensure that all violations are brought immediately to the attention of the court (i.e., swiftness and certainty of sanctioning).

The third and final dimension centers on the question of who initiated the HOPE program within each DFE site. In three of the sites, the HOPE DFE grant proposals were submitted and managed by the probation department or its parent organization, either alone or in partnership with the state court administrative office. In Clackamas County, the HOPE DFE grant was initiated outside of probation entirely. Given the importance of agency buy-in as a key facilitator of HOPE implementation, and the reality that the probation department carries the primary burden for day-to-day HOPE program operations, the probation department should play an important role in the decision to participate in a HOPE program.

4.2. HOPE DFE Implementation Fidelity

Implementation fidelity analysis was conducted for each of the four sites. Results are presented for each site and then a summarization of fidelity across the four sites is presented. As a reminder, we used two fidelity benchmarks for the 11 items examined—achieving a standard at least 60% of the time and at least 80% of the time (see Section 2.2).

Saline County, Arkansas

Exhibit 4-1 summarizes the results for Saline County, Arkansas. ***This site achieved a moderate to high level of implementation fidelity, achieving a minimum standard of 60% on nine of the eleven items. Fidelity was at 80% or greater for seven out of the eleven items.***

Fidelity was very low for Item #7 (time from a violation to a violation hearing), which provides a measure of swiftness of sanctioning, one of the primary components of the HOPE model. Only about one-third of violations (38%) were brought before the judge for a violation hearing within the 3-day time frame established as the benchmark. This may have been due at least in part to the need for the HOPE probationer officers in this site to serve as "jacks of all trades," with responsibility for most of the duties of running HOPE (e.g., offender assessment, drug testing, warrant service, regular supervision). Some or all of these tasks were performed by special purpose units within the probation offices at the other three DFE sites.

Fidelity for Item #2 was also quite weak, with most probationers admitted to HOPE having been assessed as low risk prior to enrollment in HOPE, contrary to the intended goal of admitting high risk offenders. We suspect this was due to the relatively small population of probationers available in this county, thus necessitating admission of lower risk cases. This site also had decided that HOPE eligibility would include low- and medium-risk probationers who had violations. Thus, it is possible that minimum risk cases with a violation were not reassessed when they were declared HOPE eligible because of the violation—remaining in the administrative data system with the initial minimum risk assessment although they were eligible per the local criteria. Thus, we have some reservations about characterizing Saline County HOPE implementation fidelity as unequivocally high, although they achieved at least 60%

on the other items.

Exhibit 4-1. Saline County, AR HOPE implementation fidelity

HOPE Fidelity Item	Fidelity Results
1. Leadership	83% of interviewed HOPE team members identified a leader, most commonly the HOPE Judge
2. Probationers high risk[1]	24% of HOPE probationers were moderate to high risk[1]
3. Warning hearing compliance	86% of 14 warning hearings assessed complied with the model warning hearing script
4. Initial drug testing frequency	88% of HOPE probationers had at least 8 tests in first 2 months
5. Stepped down drug testing frequency	82% of HOPE probationers had at least 1 test per month after first 2 months
6. Exceptions for Missed Drug Tests	98% of 146 cases with a missed drug test received a consequence
7. Time to Violation Hearing	38% of 639 total violations were followed by a Violation Hearing within 3 days
8. Sanction type	88% of sanctions were jail time
9. Sanction dosage	78% of jail sanctions were at or below the Hawaii HOPE mean of 19 days (Mean = 6 days)
10. Sanction certainty	97% of violations resulted in a sanction
11. Sanction swiftness	76% of sanctions began within 3 days of the Violation Hearing

[1] Data were not available for 19 cases that were counted against fidelity, as the expectation was that risk would be assessed and reported for each case. Excluding the missing cases, the valid percent assessed as moderate to high risk rises to 27%.

Essex County, Massachusetts

Essex County HOPE consisted of two separate courts that operated HOPE programs that were in many ways independent of each other although they were co-located. These two courts were Superior Court (handling more serious felony cases) and District Court (handling less serious felony and misdemeanor cases). Each court had a HOPE judge and a separate probation unit. The two courts shared a single HOPE project coordinator. Essex County was the only DFE site that had this bifurcation of its HOPE program.

Fidelity results are reported for each court separately and for the two courts combined in ***Exhibit 4-2***. ***Essex County HOPE overall achieved a moderate to high level of implementation fidelity, achieving a minimum standard of 60% on ten of the eleven assessed items. Fidelity was at 80% or greater for eight out of the eleven items.***

Like the Saline County HOPE program, Essex County HOPE struggled to achieve 60% for Item #7 (time from a violation to a violation hearing). Essex County also was low for Item #11 (time from violation hearing to the start of the sanction)—although the program was only slightly below 60% on both measures, which focus on swiftness of sanctioning, considered a critical component of the HOPE model. We speculate that problems with swiftness may have been due to significant probation office staff shortages that plagued this site throughout the DFE period, caused by a longstanding statewide hiring freeze. Looking comparatively at both courts, differences between Superior Court and District

Court were minor, and both courts met a 60% standard on the same number of items, thus supporting a picture of consistent implementation across the two courts.

Exhibit 4-2. Essex County, MA HOPE implementation fidelity

HOPE Fidelity Item	Overall Fidelity Results	Superior Court Fidelity Results	District Court Fidelity Results
1. Leadership	83% of interviewed HOPE Team members across both courts identified a leader, most commonly Superior Court HOPE Judge	88% of interviewed HOPE Team members in Superior Court identified a leader, most commonly Superior Court HOPE Judge	73% of interviewed HOPE Team members in District Court identified a leader, most commonly Superior Court HOPE Judge
2. Probationers high risk	88% of HOPE probationers across both courts were moderate to high risk	89% of HOPE probationers were moderate to high risk	87% of HOPE probationers were moderate to high risk
3. Warning hearing compliance	95% of 42 total warning hearings assessed across both courts complied with the model warning hearing script	94% of 18 warning hearings assessed complied with the model warning hearing script	96% of 24 warning hearings assessed complied with the model warning hearing script
4. Initial drug testing frequency	91% of all HOPE probationers had at least 8 tests in first 2 months	94% of HOPE probationers had at least 8 tests in first 2 months	88% of HOPE probationers had at least 8 tests in first 2 months
5. Stepped down drug testing frequency	83% of all HOPE probationers had at least 1 test per month after first 2 months	89% of HOPE probationers had at least 1 test per month after first 2 months	79% of HOPE probationers had at least 1 test per month after first 2 months
6. Exceptions for Missed Drug Tests	89% of 195 cases in both courts with a missed drug test received a consequence	88% of 69 cases with a missed drug test received a consequence	89% of 126 cases with a missed drug test received a consequence
7. Time to Violation Hearing	56% of 736 violations across both courts were followed by a Violation Hearing within 3 days	55% of 269 violations were followed by a Violation Hearing within 3 days	56% of 467 violations were followed by a Violation Hearing within 3 days
8. Sanction type	78% of sanctions across both courts were jail time	80% of sanctions were jail time	78% of sanctions were jail time
9. Sanction dosage	85% of jail sanctions across both courts were at or below the Hawaii HOPE mean of 19 days (Mean = 4 days)	72% of jail sanctions were at or below the Hawaii HOPE mean of 19 days (Mean = 4 days)	65% of jail sanctions were at or below the Hawaii HOPE mean of 19 days (Mean = 4 days)
10. Sanction certainty	91% of violations across both courts resulted in a sanction	91% of violations resulted in a sanction	92% of violations resulted in a sanction
11. Sanction swiftness	60% of sanctions across both courts began within 3 days of the Violation Hearing	60% of sanctions began within 3 days of the Violation Hearing	60% of sanctions began within 3 days of the Violation Hearing

Note: The District Court enrolled a larger number of HOPE cases than Superior Court so approximately 2/3rds of events were for district court cases.

Clackamas County, Oregon

Fidelity measures for Clackamas County are presented in ***Exhibit 4-3***. ***Clackamas County HOPE achieved a moderate to high level of implementation fidelity, achieving a minimum standard of 60% on nine of the eleven items. Fidelity was at 80% or greater for six of the eleven items.***

Fidelity was again low for Item #7 (time from a violation to a violation hearing), and Item #11 (time from violation hearing to the start of the sanction). Only about one-third of violations (37%) were brought before the judge for a violation hearing within the 3-day time frame. This may have been due to a relatively high rate of absconding within Clackamas County and the time involved in locating and apprehending the absconders given the large geographic spread of the county. In addition, some absconders were evidently never found, as inferred from the absence of a violation hearing record in the fidelity files.

With respect to Item #1 (HOPE Leadership), most team members identified someone as a HOPE leader although these nominations tended to be scattered and hesitant, especially during the intermediate and final site visits, and included the judge, probation management, probation officers, HOPE project coordinator and even the District Attorney or sheriff. This may be due to the administrative structure of probation in Clackamas County as Clackamas County Community Corrections (CCCC, probation) is part of the Sheriff's office with little direct judicial control over it.

Exhibit 4-3. Clackamas County, OR HOPE implementation fidelity

HOPE Fidelity Item	Fidelity Results
1. Leadership	92% of interviewed HOPE Team members identified a leader, no singular leader clearly identified
2. Probationers High Risk[1]	80% of HOPE probationers were moderate to high risk[1]
3. Warning Hearing Compliance[2]	40% of 5 warning hearings assessed complied with the model warning hearing script[2]
4. Initial Drug Testing Frequency	87% of HOPE probationers had at least 8 tests in first 2 months
5. Stepped Down Drug Testing Frequency	73% of HOPE probationers had at least 1 test after first 2 months
6. Exceptions for Missed Drug Tests	99% of 362 cases with a missed drug test received a consequence
7. Time to Violation Hearing	37% of 1136 total violations were followed by a Violation Hearing within 3 days
8. Sanction Type	92% of sanctions were jail time
9. Sanction Dosage	65% of jail sanctions were at or below the Hawaii HOPE mean of 19 days (Mean = 6 days)
10. Sanction Certainty	96% of violations resulted in a sanction
11. Sanction Swiftness	63% of sanctions began within 3 days of the Violation Hearing

[1] Data were not available on 37 cases that were counted against fidelity, as the expectation was that risk would be assessed and reported for each case. Excluding missing cases, the valid percent assessed as moderate to high risk rises to 98%.

[2]The on-site DFE research coordinator in Clackamas County observed a smaller number of Warning Hearings than the Coordinators in the other three DFE sites.

Tarrant County, Texas

Tarrant County, Texas HOPE achieved a very high level of implementation fidelity, achieving a minimum standard of 60% on each of the eleven items. Fidelity was at 80% or greater for ten out of the eleven items. As can be seen in ***Exhibit 4-5***, fidelity was again the lowest for Item #7 (time from a violation to a violation hearing)—although the 66% for Tarrant County was much higher than was registered in the other three sites.

Exhibit 4-4. Tarrant County, TX HOPE implementation fidelity

HOPE Fidelity Item	Fidelity Results
1. Leadership	100% of HOPE Team members interviewed identified a leader, most commonly the HOPE Judge, with some secondary endorsement of probation management and the HOPE project coordinator
2. Probationers High Risk	91% of HOPE probationers were moderate to high risk
3. Warning Hearing Compliance	100% of 19 warning hearings assessed complied with the model warning hearing script
4. Initial Drug Testing Frequency	90% of HOPE probationers had at least 8 tests in first 2 months
5. Stepped Down Drug Testing Frequency	81% of HOPE probationers had at least 1 test per month after first 2 months
6. Exceptions for Missed Drug Tests	100% of 221 cases with a missed drug test received a consequence
7. Time to Violation Hearing	66% of 1199 total violations were followed by a Violation Hearing within 3 days
8. Sanction Type	97% of sanctions were jail time
9. Sanction Dosage	93% of jail sanctions were at or below the Hawaii HOPE mean of 19 days (Mean = 5 days)
10. Sanction Certainty	>99% of violations resulted in a sanction
11. Sanction Swiftness	83% of sanctions began within 3 days of the Violation Hearing

Fidelity across the DFE Sites

Exhibit 4-5 compares implementation fidelity metrics across all four sites. The results for Essex County are shown for both courts combined as the two separate courts were substantially similar in their fidelity metrics.

The summary shows fidelity was consistent across the four DFE sites. Although Tarrant County was the only site to meet the 60% standard on all eleven fidelity items, the other sites scored at least 60% on 9 or 10 of the measures. Further, each site met an 80% standard on at least half of the items. These ***results suggest a moderate to strong degree of fidelity of implementation to the HOPE model as promulgated in the BJA solicitation.***

Exhibit 4-5. Implementation fidelity results across the DFE Sites

HOPE Fidelity Item	Saline County, AR	Essex County, MA	Clackamas County, OR	Tarrant County, TX
1. Leadership identified by HOPE team members?	83% identified a leader, most commonly the HOPE Judge	83% identified a leader, most commonly Superior Court HOPE Judge	92% identified a leader, no singular leader clearly identified	100% identified a leader, most commonly the HOPE Judge with some secondary endorsement of probation management and the HOPE project coordinator
2. Probationers High Risk	24% of HOPE probationers were moderate to high risk	88% of HOPE probationers were moderate to high risk	80% of HOPE probationers were moderate to high risk	91% of HOPE probationers were moderate to high risk
3. Warning Hearing Compliance with Model Warning Hearing Script	86% of 14 warning hearings complied	95% of 42 warning hearings complied	40% of 5 warning hearings complied	100% of 19 warning hearings complied
4. Initial Drug Testing Frequency	88% of HOPE probationers had at least 8 tests in first 2 months	91% of all HOPE probationers had at least 8 tests in first 2 months	87% of HOPE probationers had at least 8 tests in first 2 months	90% of HOPE probationers had at least 8 tests in first 2 months
5. Stepped Down Drug Testing Frequency	82% of HOPE probationers had at least 1 test per month after first 2 months	83% of all HOPE probationers had at least 1 test per month after first 2 months	73% of HOPE probationers had at least 1 test per month after first 2 months	81% of HOPE probationers had at least 1 test per month after first 2 months
6. Exceptions for Missed Drug Tests	98% of 146 cases with a missed drug test received a consequence	89% of 195 cases with a missed drug test received a consequence	99% of 362 cases with a missed drug test received a consequence	100% of 221 cases with a missed drug test received a consequence
7. Time to Violation Hearing	38% of 639 violations were followed by a violation hearing within 3 days	56% of 736 violations were followed by a violation hearing within 3 days	37% of 1136 violations were followed by a violation hearing within 3 days	66% of 1199 violations were followed by a violation hearing within 3 days
8. Sanction Type	88% of sanctions were jail	78% of sanctions were jail	92% of sanctions were jail	97% of sanctions were jail
9. Sanction Dosage: Jail days <= 19 days (Hawaii HOPE mean)	78% of jail sanctions at or below (Mean = 6 days)	85% of jail sanctions at or below (Mean = 4 days)	65% of jail sanctions at or below (Mean = 6 days)	93% of jail sanctions at or below (Mean = 5 days)
10. Sanction Certainty	97% of violations resulted in a sanction	91% of violations resulted in a sanction	96% of violations resulted in a sanction	>99% of violations resulted in a sanction
11. Sanction Swiftness	76% of sanctions began within 3 days of the violation hearing	60% of sanctions began within 3 days of the violation hearing	63% of sanctions began within 3 days of the violation hearing	83% of sanctions began within 3 days of the violation hearing
SUMMARY	**60% standard: 9 items** **80% standard: 7 items**	**60% standard: 10 items** **80% standard: 8 items**	**60% standard: 9 items** **80% standard: 6 items**	**60% standard: 11 items** **80% standard: 10 items**

All four sites did well in establishing leadership of HOPE (Item #1). More than 80% of all those interviewed identified one or more leaders of the HOPE program, most commonly suggesting leadership was vested in the local HOPE judge. While this supports HOPE as a "judge-driven" program, over time additional leaders emerged, most commonly the HOPE project coordinator and/or probation officers/management. Although those team members did not supplant the judge in the leadership role, this diffusion of leadership suggests that HOPE can operate with shared team leadership.

Three of the four sites also did well identifying moderate-to-high-risk probationers as HOPE-eligible (Item #2). The exception was Saline County, Arkansas, where less than a quarter of the HOPE probationers had a risk score of moderate or high as registered in the Arkansas probation administrative data system. We explored possible reasons for this with the local HOPE team, as well as with Arkansas probation and were unable to find a definitive answer. One possibility, which we suspect but were unable to definitively confirm, is that formal reassessments were not done when minimum risk probationers had violations and "became" HOPE eligible, which was part of the eligibility identification criteria in Saline County.

Three of the four sites also showed high compliance with the model warning hearing script developed by Judge Alm (Item #3). The model script is relatively brief and straightforward. Compliance with this item is important, as the warning hearing is one of the unique features of HOPE, where the program goals and expectations are clearly laid out for the probationers by the HOPE judge. The exception to compliance with this item was Clackamas County, where the number of warning hearing observations was low; thus, the score may not be representative of performance on this item for this site.

The four sites did quite well in following the prescribed schedule of drug testing (Items #4 and 5). Frequent random drug testing is an important component of the surveillance aspect of HOPE. The grant awards provided funds for the administration of these tests. All sites used a color-coded drug-testing hotline. Each site made its own local adaptations to cope with the workload demands imposed by the drug testing—from using existing specialized drug testing labs within probation, to hiring temporary part-time staff, to relying on the HOPE probation officers to conduct the testing. Regardless of the mechanism employed, the testing was completed well within the range established in the BJA solicitation.

The four sites also did very well with respect to the certainty of sanctioning (Items # 6 and 10). Under the HOPE model, punishment for violations must be assured, with little or no possibility for violators to negotiate their way out of a sanction or to be "cut some slack". Clearly, the four sites offered few opportunities for escape from sanctions, except for violators who absconded and were not apprehended.

Finally, ***the sites also did not over punish (Items # 8 and 9).*** Recall that it is certainty and swiftness of sanctioning that matters under the HOPE model. Severity is less important, and indeed, overly severe punishment is to be avoided. Items # 8 and 9 indicate that jail was the most common sanction type, with the typical dosage being less than 1 week, which is well within the Hawaii HOPE experience. Admittedly, these figures conceal some variation, as repeat violators were often sanctioned more heavily than first timers, but again, the punishment meted out in this DFE did not emphasize severity.

The one fidelity metric that was moderate to weak across all sites was Item # 7–the time from a violation to a violation hearing. The goal was that this timeframe would be 3 days or less. Fidelity on

this item ranged from 37% to 67%. This item measures one of the central precepts of HOPE that consequences should follow from actions as quickly as possible to reinforce the connection between actions and consequences in the minds of probationers. The violation hearing is the public stage where real consequences for misbehavior are supposed to be clearly and swiftly communicated to and imposed upon violators, thus representing a signal intending to deter future deviance. But, it was the one implementation task where all four sites seemed to struggle. As noted earlier, one reason for this was the inability of the sites to quickly find and bring in absconders.

We investigated the metrics for this measure by examining the subset of violations that were not followed by a violation hearing within 3 days to determine how far from the 3-day goal hearings were held after a violation. Results suggest that in most cases violators were brought before a hearing within a week:

- In Saline County, over 60% of violation cases exceeded the goal of holding a violation hearing within 3 days. But, the modal number of days between the violation and violation hearing for those cases that missed the desired 3-day window was 4. And, 51% of such cases had a violation hearing within 7 days. Thus, in ***Saline County, 67% of all violators were before the judge within a week.***
- In Essex County, about half of the violation cases exceeded the goal of holding a violation hearing within 3 days. But, the modal number of days between violation and violation hearing for those cases that missed the 3-day window was 5. And, 48% of such cases had a violation hearing within 7 days. Thus, in ***Essex County, 71% of all violators were before the judge within a week.*** As noted, there were few differences between the two HOPE courts in Essex County, so we report only the combined results. The mean number of days to bring a violation before the judge for those cases that exceeded the desired goal of 3 days was 21 days, reflecting outliers ranging up to 281 days.
- In Clackamas County, over 60% of violations exceeded the goal of holding a violation hearing within 3 days. The modal number of days between violation and violation hearing for those cases that missed the 3-day window was 4. And, 42% of such cases had a violation hearing within 7 days. Thus, in ***Clackamas County, 58% of all violators were before the judge within a week***. The mean number of days to bring a violation before the judge for those cases that exceeded the 3-day goal was 27 days, again reflecting outliers ranging up to 306 days. The HOPE team in Clackamas County noted that the county is very geographically large, rural and remote, which often made it challenging to track down absconders, thus contributing to this high mean.
- In Tarrant County, Texas, over 30% of violations exceed the goal of holding a violation hearing within 3 days. The modal number of days between violation and violation hearing for those cases that missed the 3-day window was 4. And, nearly 80% of such cases had a violation hearing within 7 days. Thus, in ***Tarrant County, 82% of all violations were before the judge within a week.***

We did not create a summary fidelity score across the 11 fidelity items given the uncertainties in the literature about fidelity thresholds associated with positive program outcomes. Moreover, we suspect that some items may be more important than others (such as Items # 7 and 11 which get at swiftness of

sanctioning) and thus should be weighted, but at this point we lack the empirical basis to establish such weightings.

One key question yet unresolved in the theoretical underpinnings of HOPE is exactly how swift does swift need to be? Again, the goal established by the BJA HOPE program solicitation was that a violation hearing would be held within 3 days of the violation. This goal was informed by Hawken and Kleiman (2009) who reported that approximately 70% of violation hearings occurred within 3 days of the violation in the evaluation of the original Hawaii HOPE program. But, it is unclear what if any magic is created by this 3-day period. It may suffice for the hearing to occur within a week. Or, any real deterrent effect may require that a punitive response be issued much more quickly than 3 days. For example, a hypothetical probationer may use drugs at a party on a Friday afternoon, but not actually be required to submit a urine sample until early Monday afternoon. By the time he appears before the judge for a violation hearing, it could well be Wednesday or Thursday, with him finally being placed into jail perhaps not until Friday. So, in this case, a full week would pass between the action that led to the violation (using drugs) and the full force of the consequence (jail). It is not known if this creates enough of a connection in the mind of the probationer between action and consequence, especially for offenders who are lower functioning or whose thinking is clouded by years of substance abuse.

Cook (2016:1158) argues that "…a closer look at how HOPE works in practice raises the question of the extent to which 'swift' and 'certain' apply to the timing and probability of punishment for a probation violation." He then goes on to argue that "The threat of punishment can't compete with the drug high when it comes to immediacy and, as a result, is less compelling (although still relevant). Delivering punishment within a few days rather than, say, a few months is going to increase its salience in decision making but not dramatically so." We agree with Cook's assessment and note that we simply do not have a sufficient body of evidence to support a conclusion about how long the punishment response can be delayed within a HOPE-type program without jeopardizing any deterrent effect.

The fidelity metrics presented above with respect to Item #7 (and by extension also to Item #11) may be much "better" than they look, but conversely may be much worse. Indeed, some suggest that certainty is the most salient factor in effective punishment; thus, HOPE may be able to tolerate some variation in the swiftness of consequences for violations, so long as high rates of certainty are maintained, as was the case in this DFE (Paternoster, 1987). Regardless, more research is needed on the celerity of response to better inform standards for future HOPE replications.

Implementation Fidelity Summary

The ***implementation of the HOPE DFE was a success***, based upon the analysis reported above. For most measures at most sites, implementation fidelity was moderate to high: The DFE sites did what they were supposed to in implementing HOPE. This finding should perhaps not be surprising. This DFE was monitored very closely by several sources.

- A training and technical assistance provider (Pepperdine University) was contracted to offer regular support and guidance to the four DFE sites. We learned through the course of our process evaluation that Pepperdine was a regular presence at these sites and provided ongoing feedback and correction to them about their adherence to the intended model. Moreover, the Pepperdine team also frequently involved the HOPE program originator--Judge Steven Alm—in these visits.

- While BJA delegated this program oversight to Pepperdine, staff from BJA were also periodically involved in monitoring program implementation and in offering further correction where needed.
- In addition, each site had a full-time, dedicated HOPE project coordinator whose role was to assure adherence to the HOPE principles. These project coordinators collected the fidelity metrics for the TTA provider on a real-time basis and were routinely monitored by the TTA staff.
- Finally, while the evaluation team expressly was not involved in guiding or correcting the operations of HOPE at these sites, the sites did view the evaluation team as another set of eyes watching their performance in the DFE.

There is considerable research suggesting that treatment effects are stronger when program developers or other technical assistance experts are involved closely in the implementation of a program under study (which was the case here), and most especially when the program developers *are* the evaluators (which was not the case here) (Fagan, 2013; Gandhi, Murphy-Graham, Petrosino and Chrismer, 2007; Mihalic and Irwin, 2003; O'Connell, Boat and Warner, 1994; Petrosino and Soydan, 2005). The latter, of course, raises issues about the objectivity and independence of such developer-led evaluations, but the point remains that involvement by developers in program implementation seems to enhance fidelity.

Indeed, some studies have found that early and ongoing monitoring of fidelity by experts can produce fidelity scores of over 80% (DuFrene, Noell, Gilbertson, and Duhon, 2005; Greenwood, Tapia, Abbott and Walton, 2003). Moreover, implementation sites often require multi-year technical assistance and considerable support from program experts to maximize fidelity and the odds of producing positive treatment effects (Elliott and Mihalic, 2004). Some nationally recognized program models do provide technical assistance support for sites seeking to implement their models, and in some cases the acceptance of such support is a "mandatory" condition of approval to run the program. Examples include the Nurse-Family Partnership program (Olds, Hill, O'Brian, Racine and Moritz, 2003) and others within Blueprints for Violence Prevention suite of programs (Elliott and Mihalic, 2004), and the Penn State Evidence-Based Prevention and Intervention Support Center which supplies statewide technical assistance to agencies seeking to adopt a menu of evidence-based programs funded by the state (Rhoades, Bumbarger and Moore, 2012). But, most program implementation is not guided by such expert oversight (Gandhi, et al., 2007).

The relatively strong implementation documented here for the HOPE DFE is important in that it helps to avoid a Type III error, wrongly concluding no program effects when such a conclusion may have been driven by poor implementation (i.e., there might have been a detectable treatment effect if implementation had been better).

The other take away from this DFE is that future evaluations of HOPE must carefully measure implementation fidelity, as is reported here, as well as document implementation experiences. This sort of process evaluation is critical to helping researchers, practitioners and policy makers to more fully understand the relationship between implementation and outcomes and to help them draw correct inferences about treatment effects discovered in evaluations.

4.3. Process Evaluation: Lessons Learned from the Implementation Experience

This section summarizes the cross-site qualitative results from the process evaluation of the HOPE DFE. Detailed findings for each individual site are presented in Appendix I. Topics covered include the factors that facilitated and challenged the implementation of HOPE, team members' descriptions of HOPE and PAU at their sites, HOPE team members' views about HOPE, leadership of HOPE, plans for sustainability of HOPE after completion of the DFE, and team members' perceptions of how HOPE was received and understood by the probationers. Findings were based on three rounds of site visits to the sites (see Section 2.2).

Implementation Facilitators and Challenges

The implementation of HOPE was relatively smooth and free of any significant threats to the integrity of the study. Several factors contributed to the relative ease of implementation that were common to three of the four sites:

Most notably, there was ***strong and consistent buy in to the HOPE model*** at three of the four sites. The HOPE team members at these sites believed in the value of HOPE, felt privileged to be participating in the DFE, and argued that HOPE was "the way that probation was supposed to be" and "the way that most people (who aren't involved in the criminal justice field) think that probation does operate." They saw HOPE as distinct from, and superior to, the control PAU condition. ***HOPE was praised for being consistent and fair, with a mild sanction being delivered swiftly and with certainty in response to all violations, with little room for idiosyncratic responses from individual probation officers.*** Thus, consistent with the underlying HOPE logic model, HOPE probationers know what to expect, and are held to account for their actions, but without an overly severe punishment that could be seen as unfair by the probationers.

The view of the HOPE programs was quite different from the team members' perceptions of PAU, which they viewed as essentially the inverse of HOPE – inconsistent, erratic, unpredictable and subject to the whims of the individual probationer officer. In PAU, violations would often be ignored or brushed off until a critical mass had been achieved, at which point a relatively more severe punishment would be employed, leaving probationers uncertain about what to expect from the system or which behaviors would be punished. Surveillance of probationers under PAU, in the form of drug testing and visits, was also seen to be more erratic than under HOPE. Team members' descriptions of HOPE versus PAU were remarkably common and consistent, suggesting that they were reporting on genuine features of the two approaches to probation in their jurisdictions. There was somewhat less buy-in in Oregon, where the Clackamas County Community Corrections (CCCC) operated based on the Risk-Needs-Responsivity (RNR) principles. Probation staff in Clackamas County felt that HOPE worked at cross purposes to what they believed to be more valuable evidence-based treatment (e.g., cognitive behavior therapy) and human services brokerage that they reported to have been their primary focus prior to HOPE implementation.

Another key factor promoting implementation, again in at least three of the sites, was some ***prior experience with running a HOPE-like program or substantial progress towards planning for the implementation of a HOPE-like program*** prior to the issuance of the BJA HOPE program solicitation. This was most notable in Tarrant County, Texas, where they had been operating their SWIFT program for approximately 1 year prior to the implementation of HOPE there. SWIFT is a program like HOPE that began in Texas before Hawaii HOPE. After the beginning of the HOPE DFE, SWIFT become identical to

HOPE. The HOPE team there universally cited their experience with SWIFT as being critically important to their ease in implementing HOPE. Massachusetts had been engaged in a statewide criminal justice and sentencing review and reform process for prior to the release of the BJA solicitation and had already decided to explore something like HOPE when the solicitation was released. A similar statewide criminal justice review and planning process had also been in place in Arkansas, leading to their decision to apply to be included in the DFE. Thus, participation in the HOPE DFE was not "done on a whim" in these sites, and was preceded by a period of planning and review of the needs of their systems and how HOPE might contribute to local reforms. The exception, again, was the Oregon site where the decision to apply for HOPE grant funds was made by the District Attorney with less consultation with the CCCC agency that would be responsible for implementing HOPE. Community Corrections there felt compelled to make their best efforts to implement HOPE, but indicated that they would have liked more input in the way in which HOPE was rolled out in Clackamas County.

Finally, the ***local administrative structure of probation was also important*** to the implementation of the HOPE DFE. As discussed in Zajac, Lattimore, Dawes and Winger (2015), in three of the sites (Arkansas, Massachusetts and Texas), the probation department was either directly under the control of the HOPE judge, or there was a sufficiently close administrative linkage that the judge could substantially direct the operations of probation in the service of the HOPE program. This was again most clearly noted in Texas, were probation reports directly to the judges in each county, allowing the HOPE judge to exercise considerable discretion in managing HOPE. This speaks to HOPE as a "judge-driven" model (Hawken and Kleiman, 2009), and in these three sites the HOPE judge was most commonly identified as being the leader of HOPE, although over time additional internal leaders emerged to assist the judge in guiding the program. This clear leadership facilitated the establishment of HOPE as an innovation to pre-existing probation practices. In Oregon, probation is operated locally, with the CCCC agency located in the Clackamas County Sheriff's Department and the CCCC Director reports to the Sheriff, although CCCC funding is provided by the state. The HOPE court was operated by the state court and the HOPE judge had no administrative connection to the CCCC—either direct oversight such as in Texas or contained within a common administrative structure as in Massachusetts.

Thus, ***the key factors common to the DFE sites that facilitated the replication of HOPE were strong internal support for the underlying concept of HOPE, some previous experience with a HOPE-like model, and a local administrative structure of probation that allowed for clear leadership to emerge and that resolved what otherwise could have been bureaucratic obstacles to the establishment of HOPE at these sites.***

The challenges to implementation were rather *sui generis* to each site. But, most of the sites reported that the barriers they faced were surmountable, and were simply part of doing business in a project such as this.

In Essex County, the primary challenge was the statewide hiring freeze that had been imposed on probation and all other state agencies in the wake of the 2008 financial crisis, predating the HOPE DFE. This was a "hard freeze" with few exceptions granted, and it persisted throughout the course of the DFE. Thus, unlike at the other three sites, Essex County was not able to hire special probation officers dedicated solely to HOPE, even with the receipt of the grant award from BJA. Instead, the HOPE caseload was distributed among as many as eight officers at one time, who also carried non-HOPE caseloads. Indeed, the HOPE caseloads were typically the smallest part of their overall caseload. This led

to feelings by some officers, at least early on, that HOPE was just another project that had been dumped on them without adequate resources. Beyond this, the HOPE experience in Essex County was unique, in that there were two separate courts operating HOPE – District Court and Superior Court. While none of the HOPE team members there reported this as a serious difficulty, it did somewhat complicate the implementation and evaluation, in that county. Specifically, because the two courts normally did not collaborate directly on programs, they needed to learn more about how each other operated to facilitate the rollout of HOPE there.

In Saline County, Arkansas, the primary challenge reported was ambivalence towards HOPE by the state probation administration (probation is a state function in Arkansas), although this was not reported to be a significant impediment by the local HOPE team. Considerable authority for the supervision of the HOPE probation officers had been delegated to the HOPE judge by state probation for the purposes of the DFE. Other challenges included the replacement of both HOPE probation officers midway through the DFE, and the fact that ***the officers in Saline County had to fulfill most of the major tasks included in HOPE, including drug testing, risk assessment, intake processing, warrant service and apprehension, some transportation, and of course routine supervision and case management duties***. At the other three sites, some of these tasks were delegated to specialized units. Still, none of these challenges were reported to be significant.

Finally***, in Tarrant County, Texas, the only challenge reported was the requirement for immediate arrest within the HOPE model.*** With the SWIFT court, violators were provided 24 hours to turn themselves in before a warrant was issued and this adjustment was made for HOPE in the Tarrant County.

One challenge that was common to the sites was the need to conform to the requirements of the DFE itself. The sites complained that the data collection requirements associated with the DFE were burdensome and required a lot of time from the HOPE project coordinators that each site hired with their BJA grants. In addition, the sites lamented that the probationer randomization (to HOPE and PAU conditions) requirement constituted an artificial barrier to how they otherwise would select probationers for participation, and slowed down enrollment in HOPE. Still, the sites complied very well with the terms and conditions of the evaluation, including with the randomization procedures.

Thus, the challenges faced by these three sites were relatively minor. Further, they made implementation adjustments to respond to their local needs without compromising the spirit of the HOPE model.

The one outlier to the preceding discussion of implementation facilitators and challenges was the implementation experience in Clackamas County, Oregon. As discussed in the individual site report for Oregon in Appendix I, implementation there was more troubled than at the other three sites. ***The principal issue facing implementation in Oregon was a clash of cultures between the existing probation management and supervision framework within the probation department, and the perceived demands of HOPE.*** HOPE had been brought to Clackamas County largely through the efforts of the District Attorney, with the tacit acquiescence of probation. At some risk of oversimplification, probation viewed itself as being guided by a Risk-Need-Responsivity (RNR) approach to supervision, with an emphasis on human services brokerage and utilization of evidence based treatment practices such as cognitive behavioral therapy (Andrews and Bonta, 2003). Many in probation saw HOPE as a relatively simplistic approach to supervision, with excessive emphasis on surveillance (e.g. drug testing) and

response to minor violations. They felt that HOPE detracted from the more important things that they had traditionally been doing, to the overall detriment of the probationers. This sentiment was also linked to their feeling that HOPE had been foisted on them from outside parties, without sufficient consultation. Notwithstanding these concerns, ***probation management indicated they were fully committed to making HOPE work there for the sake of the DFE and fidelity metrics bears this out.***

Other implementation challenges in Oregon included the fact that probation was administratively under the authority of the Sheriff's department, thus affording the HOPE judge less leeway in directing the management of HOPE (Zajac, et al, 2015). Sheriff's staff also cited the vast geographic spread of Clackamas County as a barrier to the swift apprehension of violators who absconded. Finally, there was significant tension between the first HOPE project coordinator and the probation department, which was resolved by the selection of new coordinator midway through the evaluation. These factors taken together made for a more difficult implementation environment than in the other three DFE sites. Despite these issues, Clackamas County, like the other three sites, did well with implementation fidelity, suggesting that they made good on their promise to give HOPE a fair shot there. ***These findings suggest that implementation can be resilient to what were some significant threats to replication of HOPE*** (Durlak and DuPre, 2008; Elliott and Mihalic, 2004; Fixsen, Naoom, Blase, Friedman, and Wallace, 2005).

4.4. Assessment of Drug Treatment Programs for HOPE Participants

While the primary implementation focus of HOPE at the DFE sites was on the delivery of swift, certain and fair sanctions in response to violations, conventional behavioral treatment, and most especially substance abuse programing, was also a part of the HOPE approach. HOPE probationers could be referred to drug treatment on an as-needed basis (after multiple failed drug tests). This treatment could include a variety of different service modalities, including lower intensity outpatient programs and longer term placements in residential facilities. It was beyond the scope of the DFE evaluation to conduct a formal evaluation of all the treatment programs employed by the DFE sites, as most sites used multiple programs and most programs were outsourced (requiring program buy-in to evaluate). During the final site visit to the DFE sites, we conducted a brief examination of the primary drug treatment program to which the *plurality* of HOPE probationer referrals was sent. Local probation officials at each of the DFE sites arranged access for us to these programs. The individual site reports included in the Appendix contain a full description of each program visited and a detailed reporting of findings. We present an overview below.

The four sites used treatment referrals to varying degrees. Tarrant County, Texas engaged treatment the most, with 56% of all HOPE probationers referred to some sort of treatment, and with 80% of all treatment referrals to the program we visited. Saline County, Arkansas referred 31% of all HOPE probationers to treatment of some sort, with 29% of all treatment referrals sent to the program we visited (the remaining treatment referrals were scattered among over a dozen other small programs). Clackamas County, Oregon referred 23% of all HOPE probationers to treatment, with 43% of referrals sent to the specific we visited. Finally, Essex County, Massachusetts referred 15% of all HOPE probationers to some sort of treatment, with 54% of referrals sent to the program we visited. Thus, across the four DFE sites, approximately one-third of HOPE probationers were referred to treatment, and we visited the four programs that served most referrals.

All four programs we visited were residential facilities that provided between 2 to 6 months of services to client. The programs we visited in Arkansas, Oregon and Texas were operated directly by either state (Arkansas and Texas) or local (Oregon) corrections agencies. The program in Massachusetts was a private non-profit provider.

The programs in Arkansas, Oregon and Texas exhibited features that would suggest they were operating in accordance with the basic principles of effective offender intervention (Andrews and Bonta, 2003; MacKenzie, 2006; MacKenzie and Zajac, 2013). They targeted criminogenic needs primarily, with less attention to non-criminogenic factors. They targeted high-risk, high-needs clients (although referrals were made by the local DFE sites, which were responsible for risk assessment for HOPE enrollees). They followed some sort of established, evidence-based treatment model (e.g., a therapeutic community or cognitive behavioral approach, among others) that included features such as manualized curricula (e.g., *Thinking for a Change*) and opportunities for behavioral practice and rehearsal. They provided clients with a structured treatment experience that kept them occupied with pro-social activities (including ancillary services such as education). They monitored client performance and advanced them through the program based on individual progress. Finally, staff were hired based on relevant training and experience, and some on-site training was provided to them. Various areas of concern at these programs included higher than desired client-staff ratios, monitoring of client behavior while in community release phases, client reward structures and absence of any formal program evaluation activities (the Clackamas County site was the only one that had undergone prior evaluation). Still, ***these three sites demonstrated features that suggest that the services they were providing could have a positive impact on clients and could be complementary to the broader sanctioning focus with the HOPE DFE.***

Finally, the program Massachusetts expressly indicated that they do not consider themselves to be a formal drug treatment program, and instead focuses on employment training and readiness and general lifestyle guidance. Thus, we were not able to examine this program as we did for the other three sites.

4.5. Communication Pathways: Network Analysis Results

Social network analysis was conducted for two measures—communication among stakeholders and assessment of involvement/importance of stakeholders. These were based on ratings reported by all stakeholders during each of the three site visits. In general, the stakeholder included HOPE judge, HOPE program coordinator, HOPE probation officers (typically two officers at each site specifically dedicated to HOPE), local probation management, representatives from the local corrections (county jail) and law enforcement (sheriff and/or local police departments) communities, district attorney (prosecution), and public defender (defense). The latter two roles were often relatively less involved in HOPE, and in the case of the HOPE program in Texas, they were not interviewed due to their absence from the operations of the program. As noted in Section 2.2, because the number of stakeholders was small, most stakeholders were linked to most other stakeholders. As a result, we looked not simply at whether there was, for example, any communication between pairs of stakeholders but at ratings of the extent of communication or involvement. Specifically, stakeholders rated their level of communication with others on a scale of 0 to 4 with anchors of “never” and “daily.” Involvement/importance was also rated on a scale of 0 to 4 with anchors of “not at all involved/important” to “very involved/important.”

The network graphs for the communication ratings are shown in ***Exhibits 4-6 through 4-10***, with separate graphs presented for the two Massachusetts courts. The network graphs for the involvement/importance ratings are shown in ***Exhibits 4-11 through 4-15***. Within each site, the upper panel has node sizes (i.e., the red circles) proportional to betweenness centrality and the lower panel has node sizes proportional to closeness centrality—the larger the red circle, the greater the betweenness or closeness centrality. If there is no circle, the value of the centrality is zero.

From left to right, each panel in each exhibit contains the network graph at each time point (baseline, interim, and final). Line width is proportional to the communication or involvement rating (or the average of the two ratings in cases of reported reciprocal communication or reported involvement). Single headed arrows indicate reports of one-way communication (e.g., the jail staff to the sheriff in the initial interview in ***Exhibit 4-6***) while double headed arrows indicate reciprocal communication (e.g., the jail staff and PC in baseline panel of ***Exhibits 4-6***). An absence of an arrow indicates no communication between stakeholders (e.g., ratings of 'never' communicating as seen between the sheriff and defense in the initial interview).

Some general patterns are immediately obvious in the exhibits. First, we consider the communication network graphs (***Exhibits 4-6 through 4-10***). The shortest path between two nodes is generally direct for the communication networks, as most nodes are connected, suggesting networks that are fully or almost fully saturated. The exception are the graphs for Texas (***Exhibit 4-10***) that show fewer direct linkages, particularly along the lower left sides of the graphs, where there are fewer direct linkages between sheriff, defense, prosecution, and probation officer. We also see that betweenness centrality often changes from interview to interview, while closeness centrality is stable across the three interviews. Second, the closeness centrality measure (i.e., the size of the red circle in the lower panel of each exhibit) is either too small to be seen or roughly the same size among those with non-trivial closeness centrality measures.

In the Arkansas site, most stakeholders had consistently low betweenness centrality except for the project coordinator (whose betweenness centrality decreased over time) and the probation officers (whose betweenness centrality increased over time). These results suggest that the project coordinator became less and the probation officer(s) became more of a conduit between other stakeholders as the DFE progressed. In the Massachusetts Superior Court there are no real patterns, although the judge, sheriff, project coordinator, and probation officers each had higher betweenness centrality during at least one-time point. A similar lack of pattern occurred in the Massachusetts District Court. In the Oregon site, the project coordinator and probation officers consistently had higher betweenness centrality that was largest at the interim interview. In the Texas site, only the project coordinator had consistently high betweenness centrality that peaked at the interim interview.

We hypothesized that as the HOPE model is a judge-driven model that the betweenness centrality should be large for the judge, but that large betweenness centrality may also be observed for other stakeholders. In fact, this is not what we observed—most likely because of the degree of saturation of the models (i.e., most individuals had direct connections to other stakeholders).

A stakeholder with a high closeness centrality score is connected to the rest of the network via shorter paths (i.e., fewer intermediary nodes) than is a stakeholder with a low closeness centrality score. In the Arkansas site, closeness centrality was low for the probation manager and sheriff, and high for all other stakeholders. In the Massachusetts Superior Court, closeness centrality was low for the sheriff,

probation manager, and prosecution, and high for all other stakeholders. The Massachusetts District Court showed some variability over time for the probation manager and sheriff (whose closeness centrality increased over time), the prosecution always had low closeness centrality, and the remaining stakeholders had consistently higher closeness centrality. In Oregon, the judge started with low closeness centrality, but had closeness centrality similar to the remaining stakeholders at the two later interviews. In the Texas site, the jail, defense, and prosecution always had low closeness centrality while the remaining stakeholders had consistently higher closeness centrality.

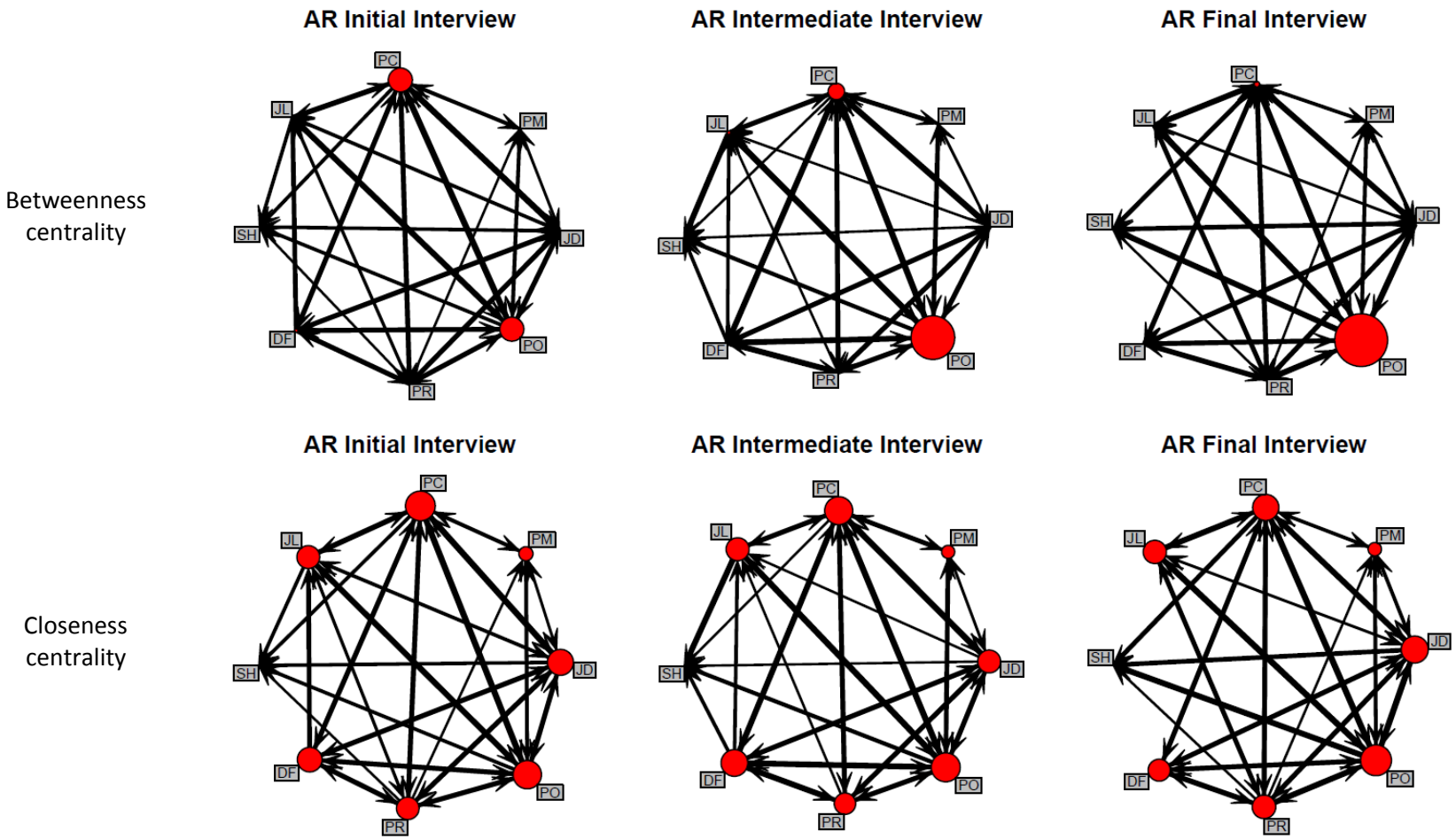
If the judge-driven HOPE model holds in practice, the closeness centrality should be large for the judge as the judge theoretically drives connections in the network. Again, this is not what we observed, with the closeness centrality of the judge similar to that of other stakeholders.

In summary, betweenness centrality measures how intermediate a stakeholder is to connections between other stakeholders, while closeness centrality measures how closely connected a stakeholder is to many other stakeholders. With the small networks considered here, most stakeholders are reciprocally connected to most other stakeholders (i.e., the networks are nearly saturated). In the context of the communication ratings, a large betweenness centrality measure indicates those who mediate the most communication exchanges about HOPE between other stakeholders, while a large closeness centrality measure indicates those who tend to communicate with the most stakeholders about HOPE.

Exhibits 4-11 through 4-15 show betweenness centrality and closeness centrality for involvement/importance ratings. In the context of these ratings, a large betweenness centrality measure indicates those who many other stakeholders rate as very involved or important to the HOPE program, while a large closeness centrality measure indicates stakeholders who tend to be highly involved with the most important of the other stakeholders. As noted earlier, prosecutors and defense attorneys played minor if any roles in HOPE in these four sites and are not included in these analyses.

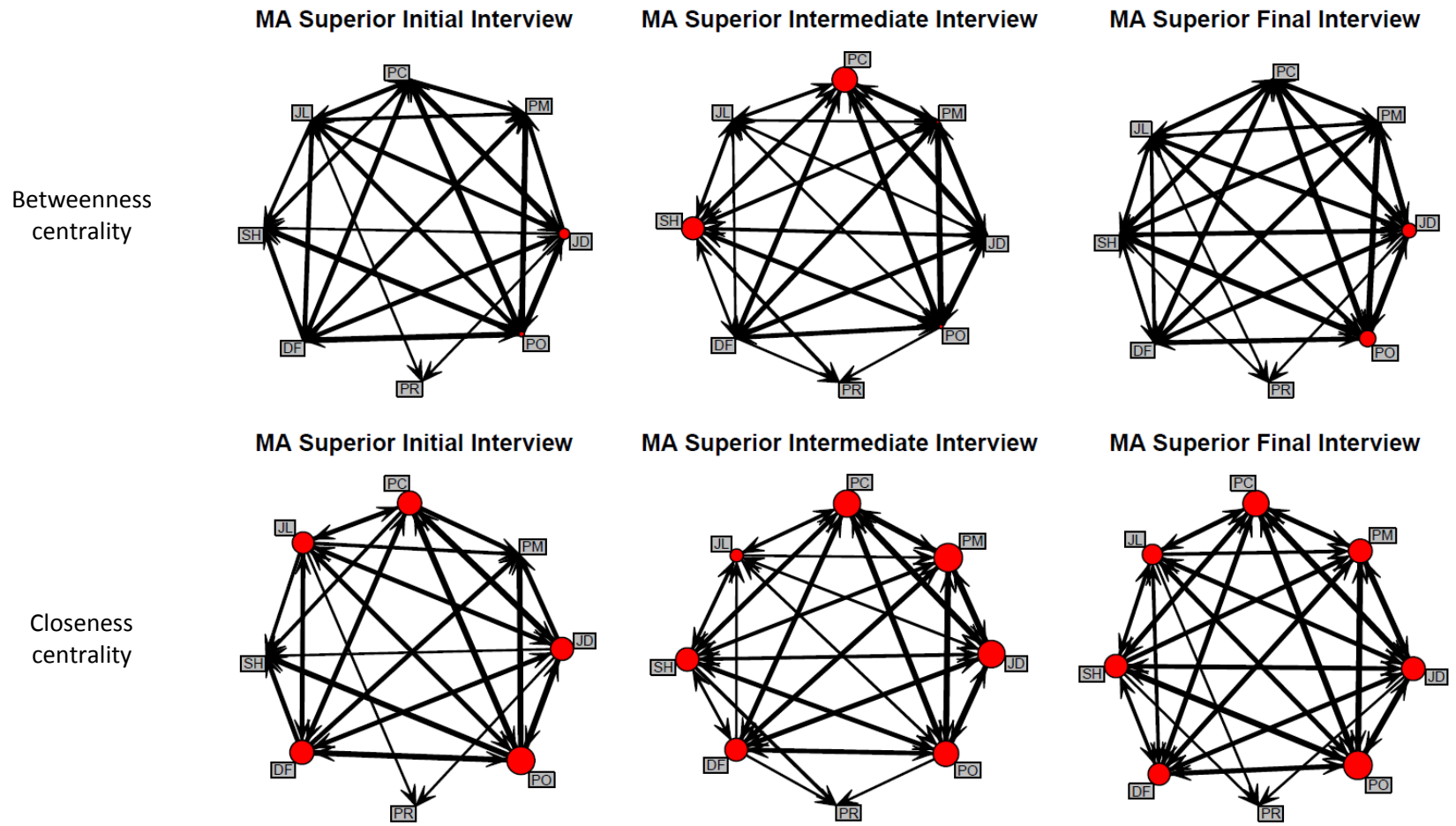
In the Arkansas site, the judge was rated as involved and important at the initial and final interviews, while the project coordinator had high betweenness centrality only at the interim interview. In the Massachusetts Superior and District Courts, only the sheriff had high betweenness centrality at the initial and final interviews. In the Oregon and Texas sites, the judge had diminishing betweenness centrality while the project coordinator had moderate betweenness centrality at the last two interviews. As with the communication ratings in the Arkansas site, closeness centrality was low for the probation manager and sheriff, and high for all other stakeholders. In the Massachusetts Superior Court, Massachusetts District Court, and the Oregon site, closeness centrality was high for all stakeholders at all time points. In Texas, only the jail staff had consistently low closeness centrality.

Exhibit 4-6. Communications network graphs for communication ratings in Saline County, AR



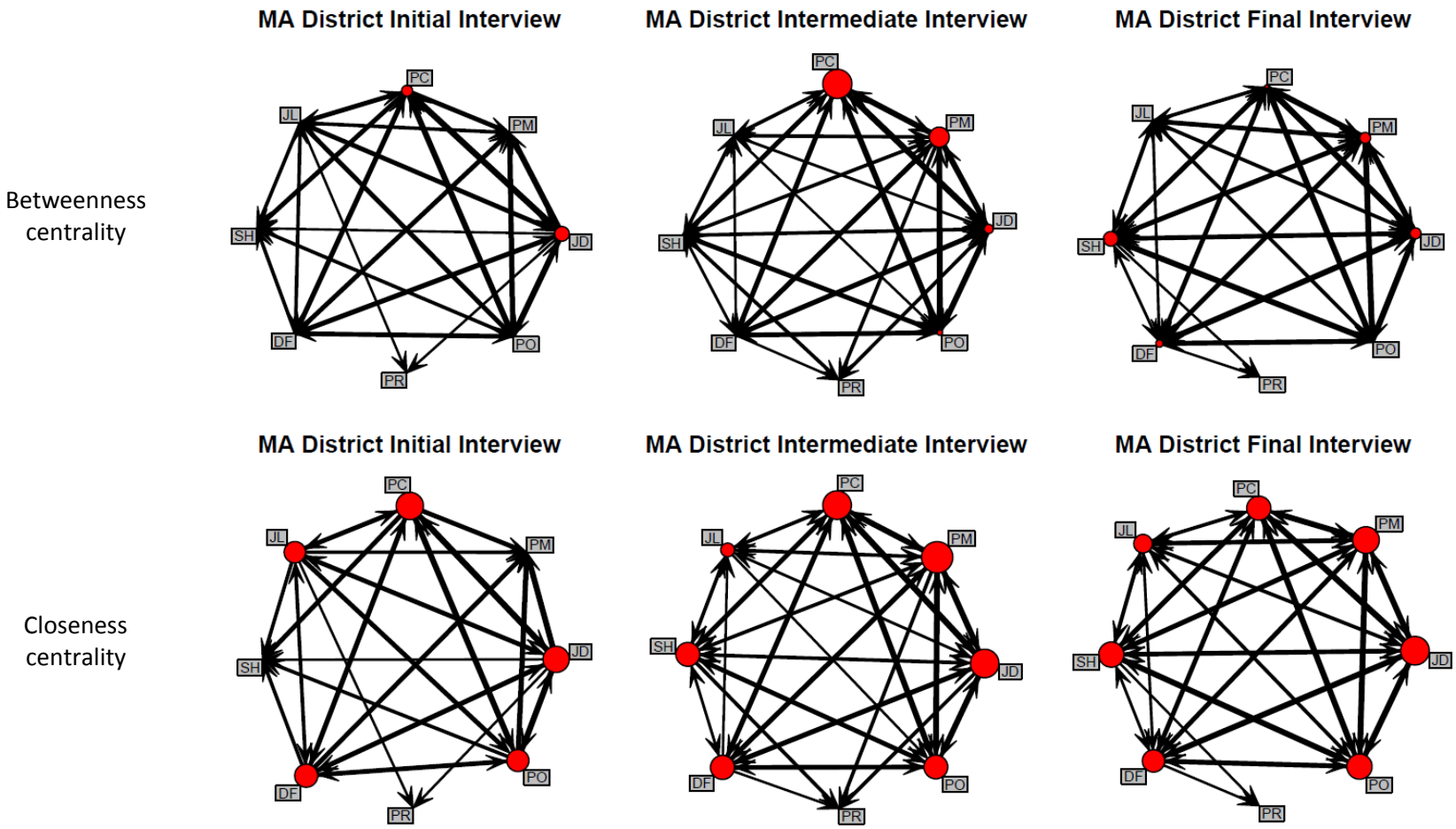
Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

Exhibit 4-7. Communications network graphs for communication ratings in Essex County, MA Superior Court



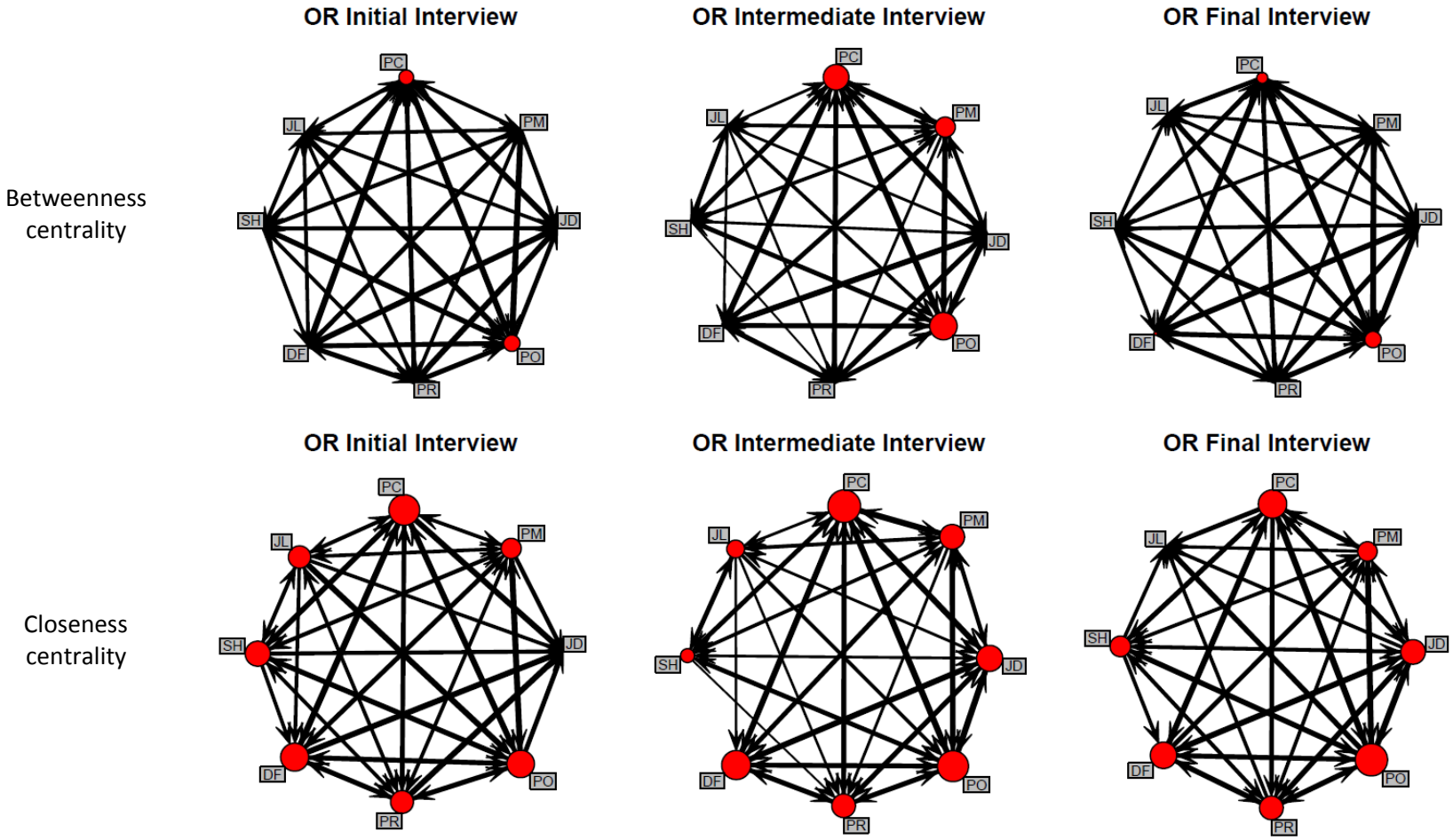
Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

Exhibit 4-8. Communications network graphs for communication ratings in Essex County, MA District Court



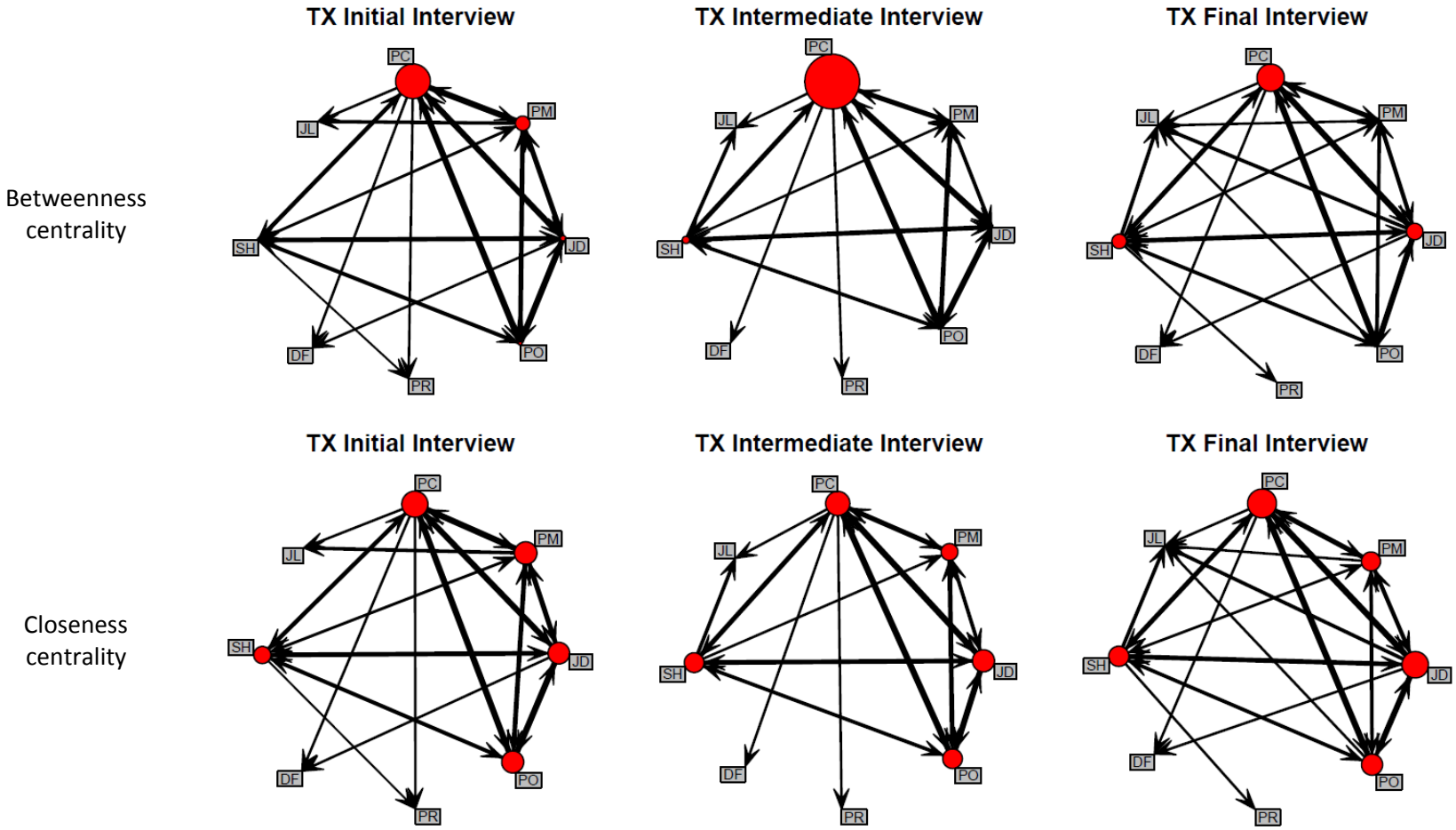
Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

Exhibit 4-9. Communications network graphs for communication ratings in Clackamas County, OR



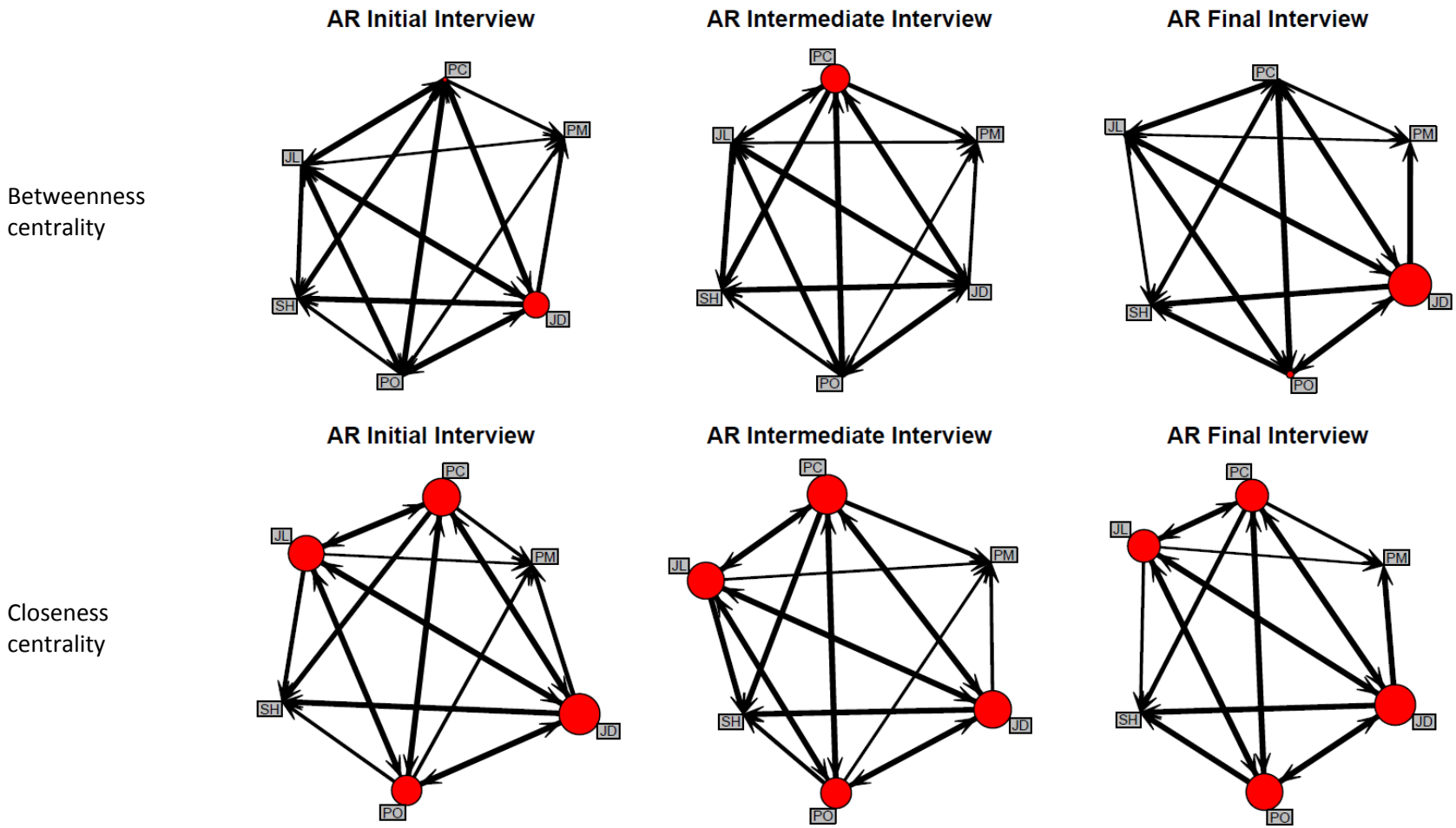
Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

Exhibit 4-10. Communications network graphs for communication ratings in Tarrant County, TX



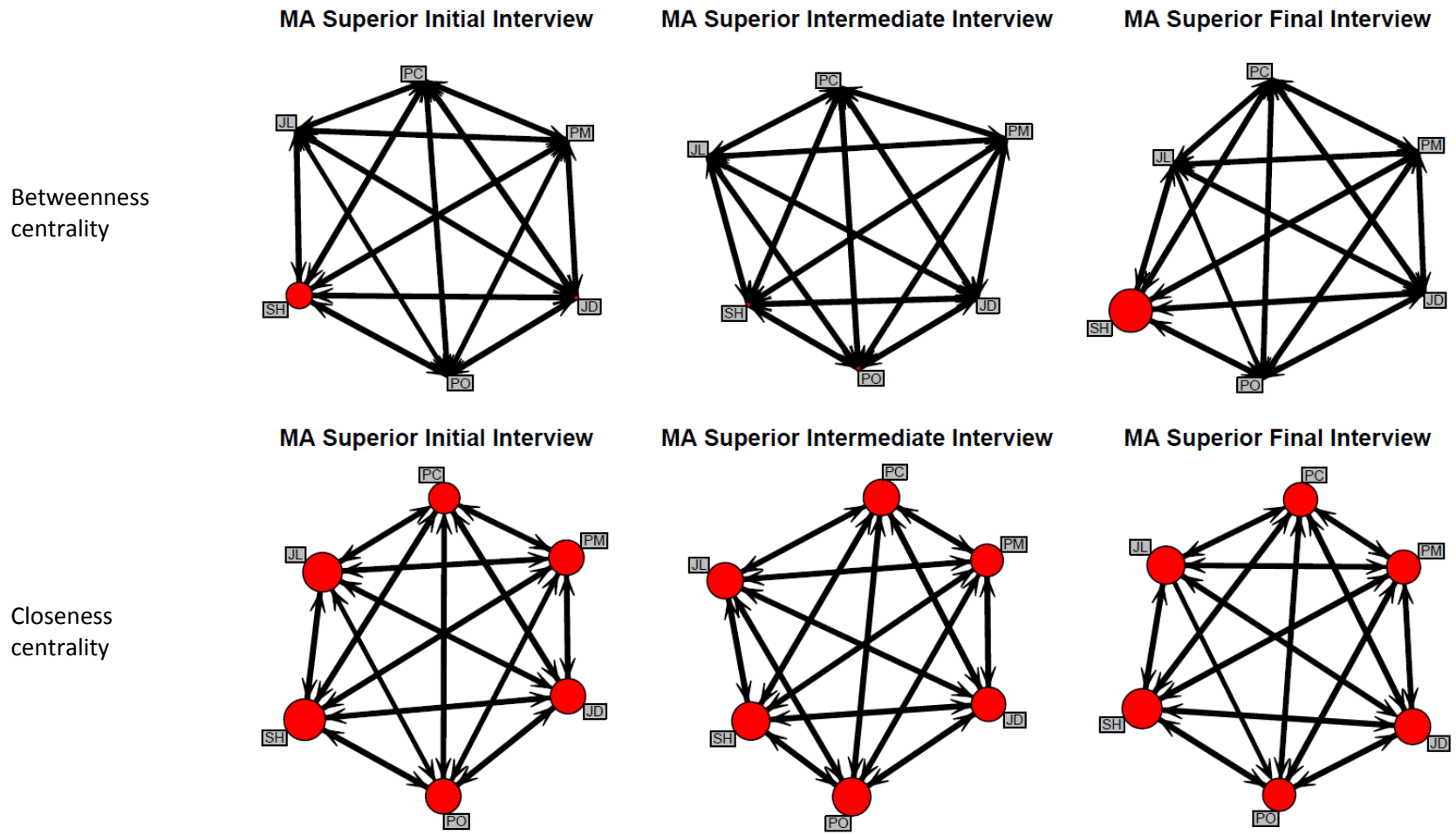
Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

Exhibit 4-11. Involvement/importance network graphs for involvement/importance ratings in Saline County, AR



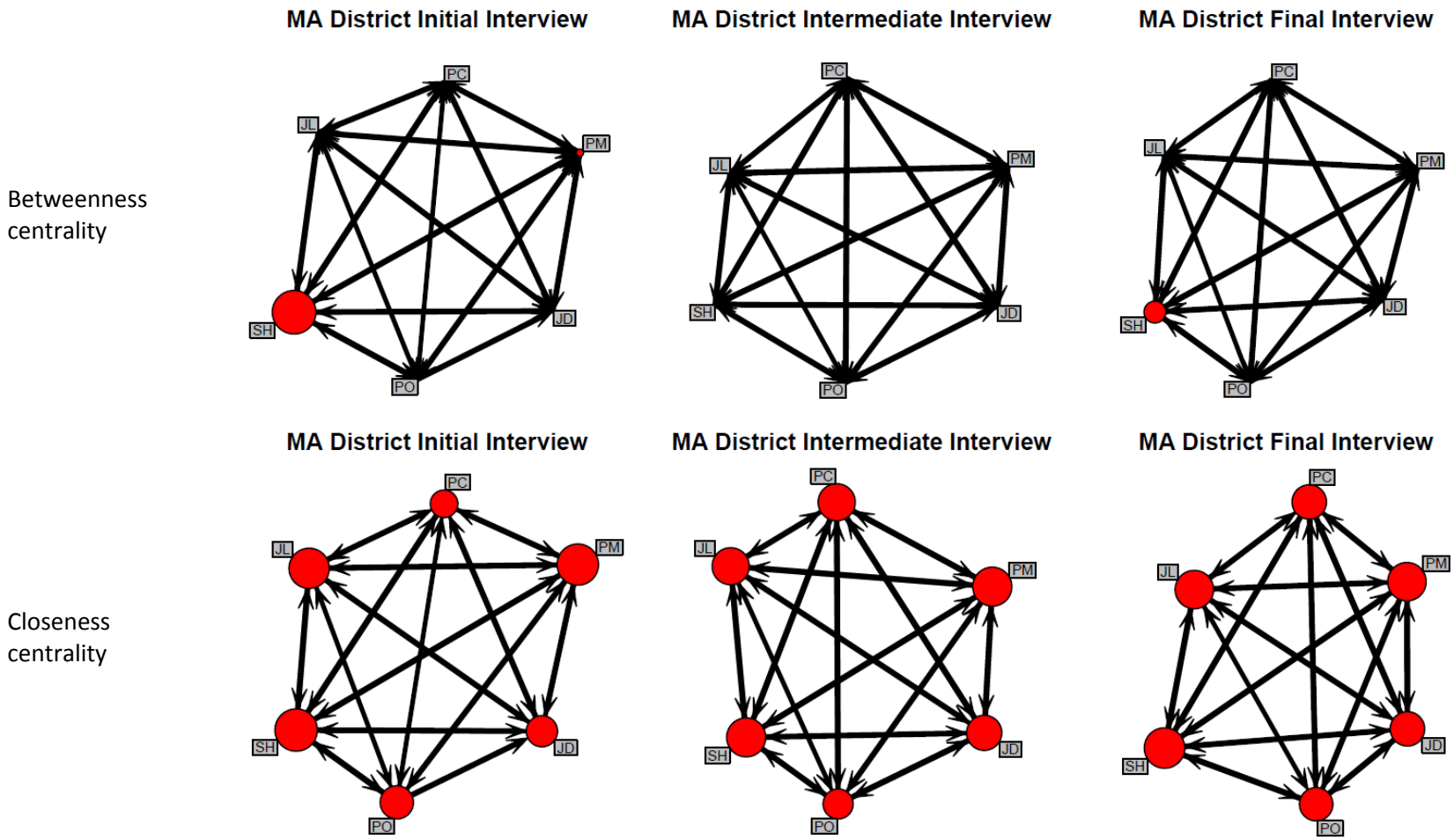
Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

Exhibit 4-12. Involvement/importance network graphs for involvement/importance ratings in Essex County, MA Superior Court



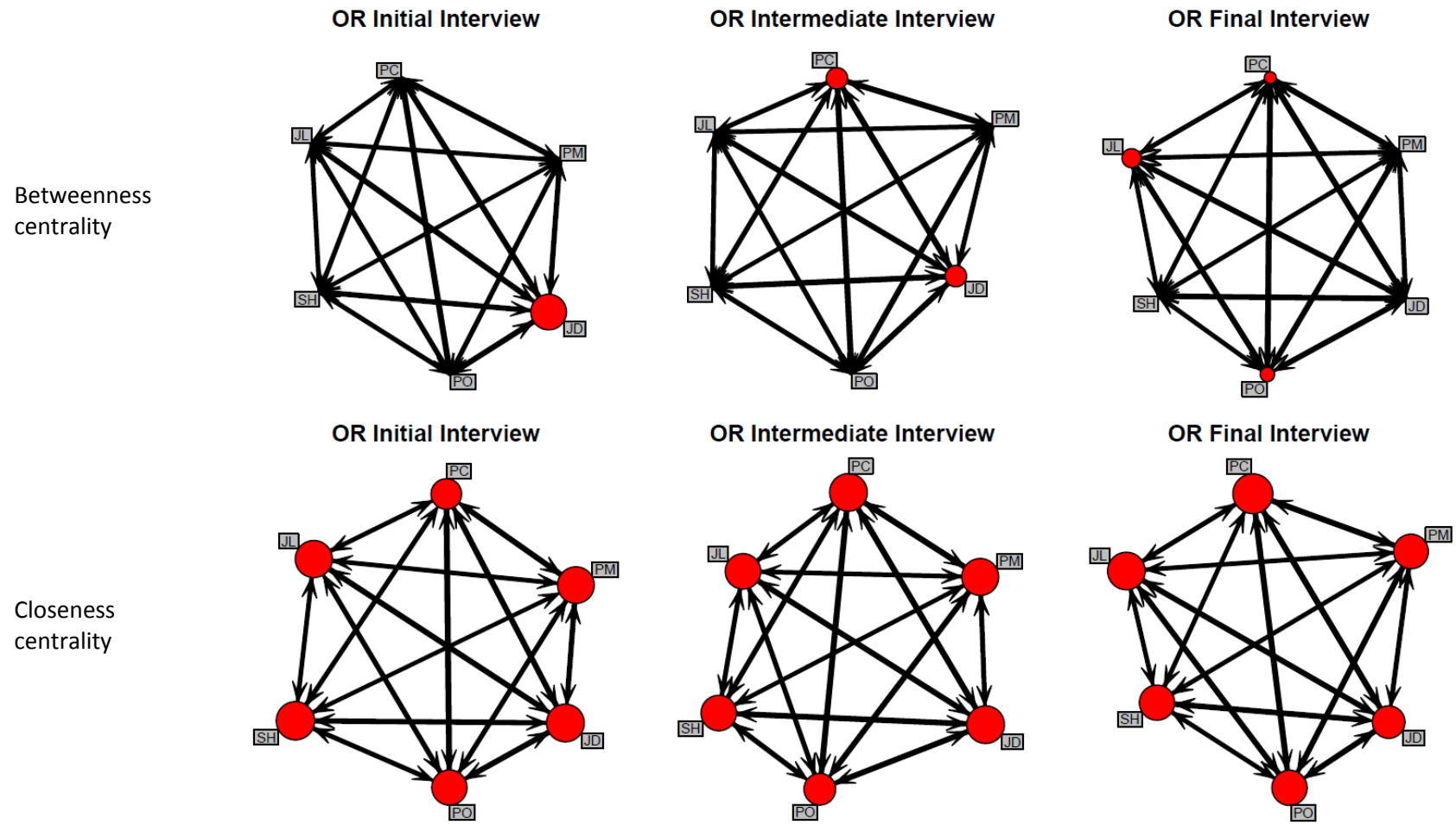
Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

Exhibit 4-13. Involvement/importance network graphs for involvement/importance ratings in Essex County, MA District Court



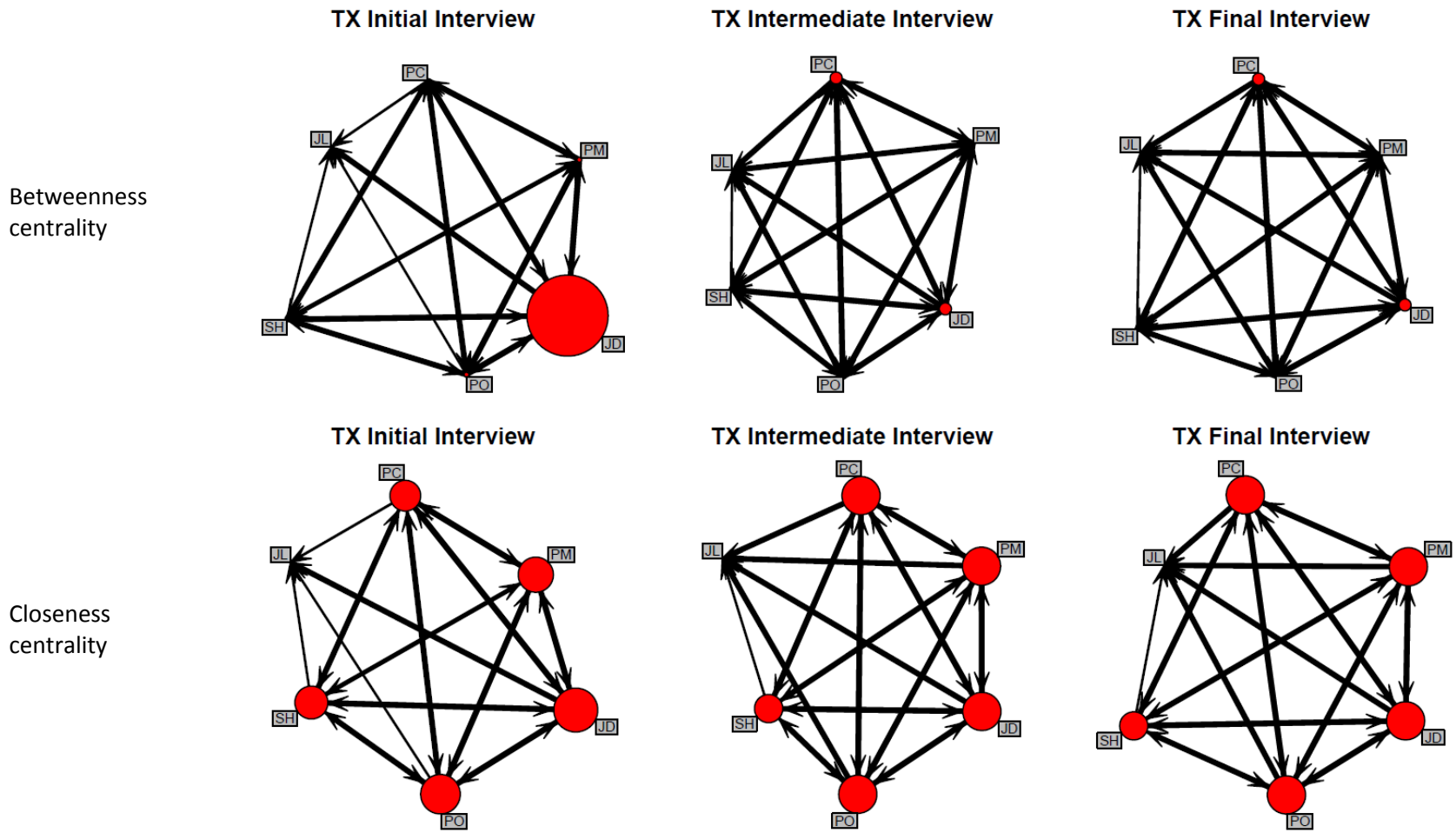
Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

Exhibit 4-14. Involvement/importance network graphs for involvement/importance ratings in Clackamas County, OR



Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

Exhibit 4-15. Involvement/importance network graphs for involvement/importance ratings in Tarrant County, TX



Notes: Node size proportional to weighted betweenness centrality (upper panel) and weighted closeness centrality (lower panel);
PC = project coordinator. PM = probation manager. JD = judge. PO = probation officer. PR = prosecution. DF = defense. SH = sheriff. JL = jail.

The network statistics density and clustering for the communications and involvement/importance measures are given in ***Exhibit 4-16*** by site and interview timing. Density is a basic descriptive measure of the saturation of ties within the network and is computed by dividing the number of ties by the total possible number of ties (i.e., maximum number of possible connections between nodes). If all nodes are tied to all other nodes, density equals 1 and there is no variance in the ties. Clustering is based on connections among triplets, rather than on pairs as is done with the density measure. A closed triplet has three stakeholders who are connected by single or double headed arrows. An open triplet includes at least one pair of stakeholders with no tie between them. There are two general types clustering measures. The first is a binary (or unweighted) measure that simply accounts for the presence or absence of communication between stakeholders. The second is a weighted measure that incorporates communication or involvement/importance ratings and takes into account the strength of connections between people. If the strength of connections (or weights) are randomly distributed in a network, the weighted measures will be equal to the binary measure. When a weighted measure is greater than the binary measure, it indicates that higher communication or involvement/importance ratings tend to appear among people in closed triplets than among those in open triplets.

In the HOPE site networks, density is very high for both communications and involvement/importance, indicating that most stakeholders communicated with each other. There was variation, however, among the sites and, to a lesser extent, over time. Oregon had the most communication among stakeholders (i.e., highest density) while Texas had the least. This difference reflects the greater amount of communication between pairs of stakeholders in Texas (saturation of the network) that is also evident in ***Exhibits 4-9*** and ***4-10***, where we see ties (lines) between all pairs of stakeholders in Oregon and many fewer connections between all stakeholders (particularly on the lower left sides of the graphs) in Texas.

Most sites had stable density over time, although both courts in Massachusetts had large increases in density between the first and second interviews. Density measures were high and stable on the involvement/importance measure across all five sites, suggesting a mutual view among stakeholders on the importance of all stakeholders.

The binary and weighted clustering coefficients are similar in each site and at each interview. This indicates that higher communication ratings in closed triplets are no more or less likely than in open triplets. In other words, three-way communication is not associated with stronger communication ratings than two-way communication. Between sites, clustering was similar in magnitude, though Massachusetts and Oregon were slightly higher (about 0.9) than Arkansas and Texas (about 0.8), suggesting the former two had stronger communication ties (and higher communication ratings) than the latter. Across time, clustering was quite stable, except in Texas which improved substantially between the second and third interviews. For the involvement/importance ratings, the network statistics often had the value of 1, indicating saturation. This is not surprising given the small, highly connected networks.

Exhibit 4-16. Network statistics for communication and involvement rating networks

	Communication			Involvement/Importance		
		Clustering			Clustering	
	Density	Binary	Weighted	Density	Binary	Weighted
AR Baseline Interview	0.73	0.84	0.87	0.67	1.00	1.00
AR Interim Interview	0.68	0.80	0.83	0.67	1.00	1.00
AR Final Interview	0.70	0.79	0.82	0.63	0.93	0.95
MA Superior Baseline Interview	0.57	0.94	0.97	1.00	1.00	1.00
MA Superior Interim Interview	0.77	0.90	0.92	1.00	1.00	1.00
MA Superior Final Interview	0.77	0.91	0.93	0.97	0.96	0.97
MA District Baseline Interview	0.55	0.92	0.94	0.97	0.96	0.97
MA District Interim Interview	0.79	0.90	0.92	1.00	1.00	1.00
MA District Final Interview	0.79	0.95	0.97	1.00	1.00	1.00
OR Baseline Interview	0.82	0.95	0.96	1.00	1.00	1.00
OR Interim Interview	0.88	0.90	0.91	1.00	1.00	1.00
OR Final Interview	0.80	0.91	0.91	1.00	1.00	1.00
TX Baseline Interview	0.46	0.79	0.85	0.80	0.95	0.96
TX Interim Interview	0.41	0.77	0.82	0.83	1.00	1.00
TX Final Interview	0.50	0.89	0.92	0.83	1.00	1.00

Among the network characteristics of the left panel of ***Exhibit 2-12***, the characteristic leading to the best fitting exponential random graph model (ERGM) for each network at each interview is indicated in ***Exhibit 4-17***. Best fit is for all possible triplets among all stakeholders within each site and interview.

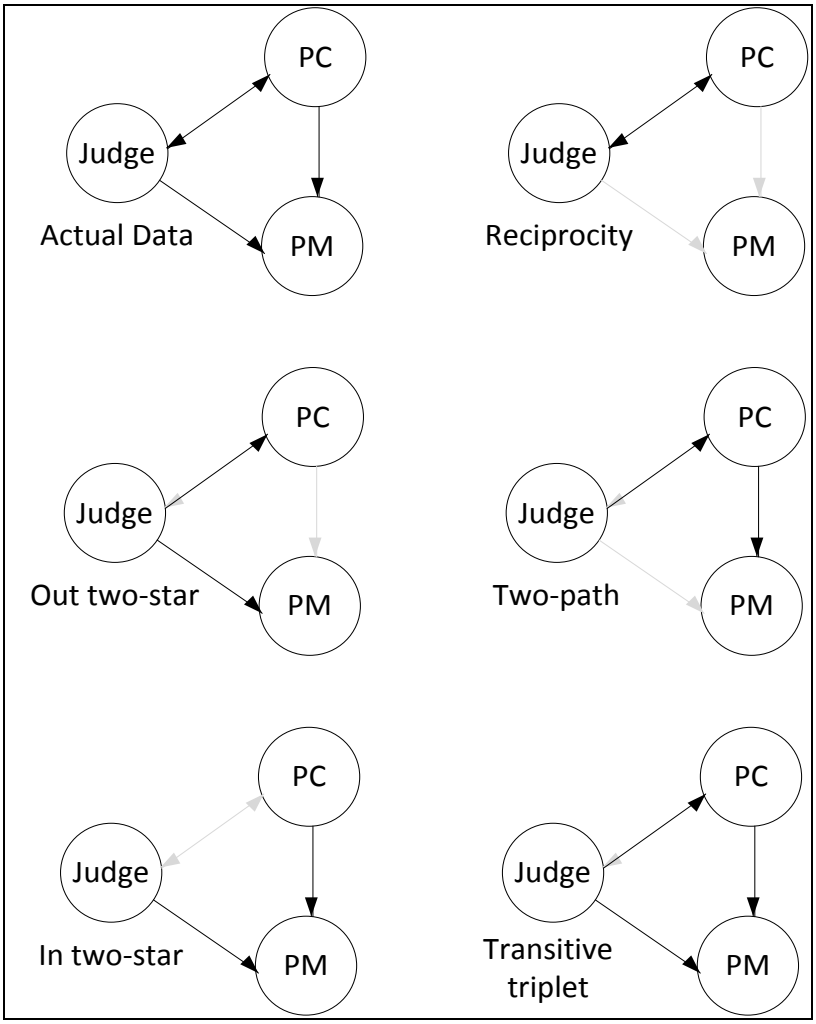
For communication ratings, the out two-star structure most frequently best described the observed ties between stakeholders. The Arkansas site had an out two-star structure as the network characteristic that best described the ties between stakeholders at all three interviews and for both measures (communication and involvement/importance). This structure is shown for the judge, project coordinator, and probation management for the baseline Arkansas data in ***Exhibit 4-18***. The actual, out two-star structure among the triplets is shown in the upper left. The Massachusetts District Court and Oregon sites also had out two-star as the best fitting structure in two of three interviews communications. The Massachusetts Superior Court started with an out-two-star structure but transitioned to the transitive triplet for the two later interviews. Since the transitive triplet is an out-two-star structure augmented with an additional tie between the two receiving stakeholders, this evolution suggests that communication relationship became more connected as time went on. The Texas site also improved from simpler communication ties to the transitive triplet as time went on. Thus, for these sites and measures, the best description of the ties is that information is being sourced from one stakeholder to multiple other stakeholders.

Exhibit 4-17. Best-fit network structure for communications and involvement by site

Site Visit Findings	Communication Best Fit Network	Involvement/Importance Best Fit Network
AR Baseline Interview	Out two-star	Out two-star
AR Intermediate Interview	Out two-star	Out two-star
AR Final Interview	Out two-star	Out two-star
MA Superior Baseline Interview	Out two-star	In two-star
MA Superior Intermediate Interview	Transitive triplet	In two-star
MA Superior Final Interview	Transitive triplet	No structure
MA District Baseline Interview	Out two-star	No structure
MA District Intermediate Interview	Out two-star	In two-star
MA District Final Interview	Reciprocity	In two-star
OR Baseline Interview	Out two-star	In two-star
OR Intermediate Interview	In two-star	In two-star
OR Final Interview	Out two-star	In two-star
TX Baseline Interview	Reciprocity	Out two-star
TX Intermediate Interview	No structure	Out two-star
TX Final Interview	Transitive triplet	Out two-star

Alternative possible structural arrangements are shown in the remaining five graphs in ***Exhibit 4-18***. For the involvement/importance ratings, the out two-star structure consistently best described the observed network in the Arkansas and Texas sites. Again, this suggests a network where one stakeholder perceived importance in other stake holders. In contrast, the in two-star structure (bottom left graph in ***Exhibit 4-18***) was most common in the remaining three sites, suggesting that other stakeholders perceived importance in a common stakeholder.

Exhibit 4-18. Results from initial interviews for the AR site for three stakeholders



Note: Structures shown as special cases by greying out components of the actual data.

4.6. Stakeholder and Probationer Perspectives' of HOPE

Interviews with HOPE stakeholders and probationers, as well as survey data that assessed probationers' attitudes over time, provide insight into the impressions and impacts of HOPE.

HOPE Team Members' Views of HOPE

Generally, HOPE team members were very positive about the concept of HOPE and its application in their jurisdictions. They believed that HOPE was "the way probation should be," and the "future of probation"; some felt "liberated" (from hidebound probation practices) by HOPE. They were optimistic that HOPE would be successful, but were cautiously awaiting the results of the evaluation before drawing definitive conclusions about HOPE. HOPE seemed to provide to them the promise of a better way forward for probation. These feelings were especially notable as the DFE wore on, with team members who expressed some skepticism early on having in the words of one team member "drunk the Kool Aid" by the end of the DFE.

HOPE team members were asked how they thought probationers received the HOPE program and what impacts participating in HOPE probation may have had. Specifically, team members were asked:

1. How well they thought the probationers understood the goals and operations of HOPE and what was expected of them under HOPE;
2. How well they thought the sanctioning regime under HOPE would change the criminal behavior of probationers (in advance of any outcome findings); and
3. Any other general impressions of how HOPE impacted probationers, including the team member's views about which components of HOPE worked well and which components did not work well.

The HOPE teams reported that their probationers understood HOPE well, with responses on a scale of 1 to 5 (low to high) ranging from 3.9 in Massachusetts (no difference between the two HOPE courts) to 4.4 in Texas. The HOPE teams reported that the warning hearings provided a good vehicle for explaining HOPE to the probationers and for resolving their uncertainties. Warning hearings witnessed by the evaluation team confirmed that probationers would ask questions and receive clarification on HOPE expectations. Team members also indicated that the HOPE probation officers, and in some cases the prosecutors and defense counsel, played an important role in explaining HOPE to probationers, even in advance of the warning hearings. Thus, it appears that the HOPE probationers were thoroughly briefed on the requirements of HOPE.

POSITIVE VIEWS OF HOPE

HOPE team members had positive views of HOPE believing it was the way probation should be. Further, the team members stated that they thought HOPE was helping probationers learn to better manage their lives.

The ***HOPE team members also indicated that they thought that HOPE is effective in changing probationers' behavior.*** Ratings on the 1-to-5 scale ranged from 3.5 in Oregon to 4.6 in Arkansas. Thus, the HOPE teams had high expectations that HOPE would reduce criminal behavior of enrolled probationers and increase pro-social behavior.

The stakeholders also indicated that the ***implementation and operation of HOPE was little burden to them***. Team members were asked to rate on a scale of 1 to 5 (low to high) the extent to which HOPE was a burden to them personally, and to other departments that were participating in the DFE. Mean ratings of burden were low, ranging from 1.7 in Arkansas to 2.8 in Massachusetts. Perhaps not surprisingly, respondents tended to indicate that the greatest burden fell on probation, and specifically on the HOPE probation officers; but, no one indicated feeling overwhelmed by HOPE. Further, respondents thought any burden was "worth it." The burden may have been eased by the relative stability of the HOPE teams throughout the DFE, with little turnover and with consistent patterns of involvement and communication between the team members.

When asked which aspects of HOPE they thought worked best for the HOPE probationers, the HOPE ***team members principally echoed their own observations about what they themselves liked about HOPE – consistency, fairness, the setting of clear behavioral expectations, the message to probationers that "we mean what we say," and the intensive surveillance including frequent drug testing.*** They felt that these elements were the most helpful to probationers in changing their life

course. Another aspect that emerged strongly from the HOPE team members was ***the belief that HOPE was helping probationers to learn how to better manage their lives more generally***. The setting and enforcement of clear expectations within HOPE was thought to teach participants time and task management, discipline with finances, timeliness, diligence at work, and overall life skills management. Because HOPE required them to be on time (to the minute) with drug testing, probationer officer visits, hearings before the judge and the like, the resulting personal discipline gained through HOPE was thought to translate to other areas where they had been challenged by dissolute lifestyles. As discussed further below, probationers who felt that HOPE had helped them also cited this life management benefit of the program. Thus, the benefits of HOPE, from the perspectives of both the HOPE team and the HOPE probationers, seemed to focus on specific changes to the traditional probation supervision model, as well as more general spill-over benefits to how probationers were living their lives because of HOPE.

At something of a cross purpose to this perceived impact on improving life skills, the HOPE teams ***also frequently noted that the intensive drug testing regimen under HOPE could create conflicts for probationers who were employed, especially in the construction industries***. These probationers often needed to be at work when the drug testing centers were open, perhaps at job sites distant from the centers, making it challenging to present themselves for required testing. Their employers also were not always sympathetic to their need to submit drug tests, caring only whether the employee showed up for work on time. The HOPE teams would make allowances for this to the extent possible, such as extending testing hours or even reaching out to employers to ask for consideration, but they acknowledge that there was only so much they could do, and that this is an inherent tension within the HOPE model. This problem was also noted by the probationers themselves and is discussed further in the chapter reporting on the qualitative interviews with the probationers.

There was overall a high degree of satisfaction with the HOPE experience among the HOPE teams at most of the DFE sites. Apart from the Oregon DFE site, few negative consequences were reported for HOPE. Still, HOPE team members offered some thoughts about features of HOPE that were <u>less</u> desirable and what they saw as limitations of the model:

1. While the HOPE model is predicated on the ***consistent*** enforcement of standards of behavior and application of sanctions for violations of those standards, some HOPE team members lamented the lack of discretion that the probation officers had in HOPE compared to PAU—reflecting tension between a zero-tolerance approach to probation and an approach that takes the circumstances of each case into consideration. This conflict was discussed repeatedly by various HOPE team members.
2. There was also concern expressed that HOPE was not working as well for the more seriously antisocial probationers, who were not impressed by the prospects of serving jail time. These probationers were reported to have a cavalier attitude towards HOPE and were inclined to do what they wanted, regardless of the consequences. For example, a probationer who missed a drug test because he was with his girlfriend stated that what he was doing at that time was far more important than compliance with the HOPE mandate and he was more than happy to take the jail time (as he had been in jail many times and was not bothered by the prospect). Another probationer at a violation hearing witnessed by the evaluation team was openly contemptuous of the process, laughing at the judge and

quite pleased to be making a spectacle of himself. These are anecdotes, but illustrative of impressions related to us by the HOPE teams. In our interviews with HOPE probationers, reported later, we also picked up on a dichotomy between probationers who were being reached by the HOPE message, and those who were not.

3. There was also concern that the "HOPE message" did not register with lower functioning and mentally ill probationers. These probationers were thought to need greater structure and support, with one HOPE judge arguing for a separate mental health HOPE court where the HOPE team could work more closely with these probationers to help them better process the requirements of HOPE.
4. Some HOPE team members also relayed concerns that drug treatment was downplayed in the HOPE model, despite the discussion among HOPE proponents of "behavioral triage". For the seriously addicted probationers, these team members believed that intensive treatment was even more important than the focus on swift and certain sanctioning, a view echoed in our interviews with the HOPE probationers (e.g. "serious addicts need treatment, it's unrealistic to expect HOPE itself to do much for them"). The HOPE teams were quick to point out that they provided treatment where they felt it was needed, even where they felt that HOPE proponents would view treatment as a substitute for a jail sanction.
5. Another concern was that for some probationers, HOPE itself had become something of an addiction. Some HOPE probationers asked to be kept on the drug testing "color line" even after it was no longer required and others asked if they could continue to come in for drug tests even after they had completed their terms of probation. We also heard this in our probationer interviews. These probationers were afraid that once they were no longer being held accountable by the intensive surveillance regimen of HOPE that they would relapse to drug use or to other undesirable behaviors. Thus, they felt they needed the structure of HOPE to maintain their good behavior. Similar findings were noted in the original Hawaii HOPE evaluation (Hawken and Kleiman, 2009). Of course, the surveillance cord must be cut eventually. It is encouraging that these probationers embraced the oversight provided by HOPE and saw benefit to it, but at the same time this leaves us uncertain what will become of them once their supervision ends.

Overall, HOPE team members saw few major downsides to HOPE. The preceding concerns can be considered notes at the margins of the DFE. Team members were consistently optimistic that HOPE would produce good results for their clients.

HOPE Probationers' Perspectives

Information on HOPE probationers' perspectives on HOPE were available from two sources: (1) the ACASI interviews conducted prior to randomization (Wave 1), and at 6 months (Wave 2) and 12 months (Wave 3) after randomization, and (2) qualitative interviews conducted with 21 HOPE probationers during the final process evaluation site visits in the fall of 2014 that provide greater detail about the experience of HOPE probation.[29]

[29] Given the relatively small number of interviews at each individual site, we report only cross-site findings and do not draw comparisons between sites. Probationers were selected by the local RTI research coordinators, and we required that the interviewees had been in HOPE for at least 6 months. For most, their time in program was

Sample Demographics

As described in Section 3, HOPE probationers who participated in a Wave 1 ACASI interview averaged nearly 31 years old and PAU probationers averaged 32 years old (***Exhibit 3-13***). About 80% of HOPE and PAU probationers were male, nearly 70% were white, and about 15% were black. About 15% of HOPE and PAU probationers were of Hispanic origin. Nearly 34% of HOPE probationers and 31% of PAU probationers reported being currently employed. There were no significant differences in these characteristics between HOPE and PAU probationers at the Wave 2 or Wave 3 interviews.

The sample of 21 HOPE probationers we interviewed included 12 men and 9 women. Most were ages 20-40 years (N=14), with the remainder older, although none was older than late 50s. Most were white (N=16), consistent with the overall demographics of the HOPE sample. Approximately one third reported that they were regularly employed, often in construction or service occupations, with the remainder claiming to be unemployed or whose status was unclear or shifting.

Perceptions about the Consequences of Probation Noncompliance

A necessary, if not sufficient, condition for HOPE to deter probation violations is for the probationer to understand what conditions would result in a violation and to believe that noncompliance would result in a sanction. At each interview wave, HOPE and PAU probationers were asked whether they thought specific results would follow if they did not comply with their probation conditions, including if they admitted to using illegal drugs, tested positive for illegal drugs, skipped drug tests, missed appointments with their probation officer, or failed to attend required treatment. Probationers were asked to indicate on a scale of 1 to 9 how strongly they agreed with 10 statements about specific responses to noncompliance (***Exhibit 4-19***). Responses to these items were summed to create a "deterrence score" for each probationer that ranges from 10 to 90, with a higher score indicating greater deterrence.

At the Wave 1 interview, HOPE and PAU probationers shared similar perceptions about the consequences of noncompliance. First, ***nearly all HOPE and PAU probationers reported that they understood what behaviors would result in a probation violation (96% and 95%).*** On average, HOPE and PAU probationers reported high levels of agreement that their probation officers would find out about their noncompliance and would arrest them or have them arrested for noncompliance. Both groups generally disagreed that the judge would do nothing in response to noncompliance. HOPE and PAU probationers expressed higher average levels of agreement that the judge would impose a week or more in jail (6.79 and 6.49), compared to their perception that the judge would impose less than a week in jail (2.48 and 2.51). On average, both groups also expressed relatively high levels of agreement that the judge would revoke their probation for noncompliance (7.05 for both groups). Similar findings were found when probationers were asked about the consequences of failing to comply multiple times.

closer to 1 year and many were close to completing HOPE. In the narrative, we provide general examples related to us, where such examples are not so specific to an individual that it would jeopardize confidentiality.

Exhibit 4-19. Self-reported responses to noncompliance by group and wave

Question	Wave 1 means		Wave 2 means		Wave 3 means	
	HOPE	PAU	HOPE	PAU	HOPE	PAU
Overall, do you think you understand what behaviors will result in a violation of probation?	96.3%	94.9%	95.7%	95.7%	97.4%	95.3%
My probation officer would find out if I did not comply with the conditions of my probation. (1=Strongly disagree to 9=Strongly agree)	7.97	7.73	**8.31***	**7.83**	8.13	7.71
My probation officer would arrest me or have me arrested if I did not comply with the conditions of my probation. (1=Strongly disagree to 9=Strongly agree)	8.07	7.96	**8.61****	**8.09**	8.34	8.03
The judge would do nothing if I did not comply with the conditions of my probation. (1=Strongly agree to 9=Strongly disagree)	8.04	8.00	8.36	8.27	**8.50****	**7.87**
The judge would give me less than a week in jail if I did not comply with the conditions of my probation. (1=Strongly disagree to 9=Strongly agree)	2.48	2.51	**3.16****	**2.61**	3.25	2.74
The judge would give me a week or more in jail if I did not comply with the conditions of my probation. (1=Strongly disagree to 9=Strongly agree)	6.79	6.49	6.77	6.58	7.02	6.73
The judge would revoke my probation if I did not comply with the conditions of my probation. (1=Strongly disagree to 9=Strongly agree)	7.05	7.05	**6.32****	**7.14**	6.78	7.36
The judge would do nothing if I did not comply with the conditions of my probation multiple times. (1=Strongly agree to 9=Strongly disagree)	7.90	8.15	8.28	8.07	8.26	7.84
The judge would give me less than a week in jail if I did not comply with the conditions of my probation multiple times. (1=Strongly disagree to 9=Strongly agree)	2.10	2.06	1.92	2.23	1.97	1.93
The judge would give me a week or more in jail if I did not comply with the conditions of my probation multiple times. (1=Strongly disagree to 9=Strongly agree)	6.78	6.49	7.09	6.58	**7.26****	**6.29**
The judge would revoke my probation and send me to prison if I did not comply with the conditions of my probation multiple times. (1=Strongly disagree to 9=Strongly agree)	7.58	7.58	7.49	7.55	7.48	7.35
Deterrence score (10-90; higher score=greater deterrence)	64.61	63.97	66.34	64.89	**67.10****	**63.72**

** HOPE and PAU differ at $p < 0.01$; * HOPE and PAU differ at $p < 0.05$

Significant differences in expectations between the HOPE and PAU respondents emerged at the Wave 2 and Wave 3 interviews—presumably a result of having been exposed to HOPE. At Wave 2, HOPE probationers expressed a higher level of agreement than PAU probationers that their probation officer would find out about their noncompliance (8.31 versus 7.83), and that their probation officer would arrest them or have them arrested for noncompliance (8.61 versus 8.09). PAU probationers expressed a higher level of agreement than HOPE probationers that the judge would revoke their probation for noncompliance (7.14 versus 6.32)[30]. HOPE probationers expressed a low, but higher level of agreement than PAU probationers that the judge would impose less than a week in jail for noncompliance (3.16 versus 2.61). ***The level of disagreement expressed by HOPE probationers that the judge would do nothing in response to noncompliance grew higher over the interview waves, and was significantly higher than PAU probationers at the Wave 3 interview (8.50 versus 7.87)***. Additionally, the level of agreement expressed by HOPE probationers that the judge would impose a week or more in jail in response to multiple noncompliance events grew higher over the interview waves, and was significantly higher than PAU probationers at the Wave 3 interview (7.26 versus 6.29). Across all interview waves, both groups reported consistently high levels of agreement that the judge would revoke their probation in response to multiple noncompliance events. Finally, ***the average deterrence score among HOPE probationers grew higher (i.e., greater deterrence) over the interview waves, and was significantly higher than PAU in the Wave 3 interview (67.10 versus 63.72)***.

CHANGING EXPECTATIONS

At baseline, HOPE and PAU probationers had similar expectations about the consequences of complying with their supervision conditions. The HOPE group expressed greater expectations of consequences over the follow up period, while expectations remained about the same as baseline for the PAU group.

Through qualitative interviews, we explored whether the HOPE probationers understood the goals, terms and conditions of HOPE as these applied to them. Rather than simply asking them if they understood HOPE (which could produce socially desirable responding), we asked them to describe HOPE to us and to explain in their own words what was expected of them under HOPE (see interview instrument in Appendix B). ***All 21 respondents were able to offer cogent explanations of HOPE, demonstrating their understanding of key features and expectations***, such as the need to attend any and all hearings before the HOPE judge, to call into the drug testing color line on schedule, to report for drug testing when required by the color line, to report for meetings with their probation officer, to adhere to other special terms and conditions imposed on them (e.g., employment), and to be honest with the judge and their probation officers about any mistake they made. They related that they understood that HOPE would hold them accountable for all mistakes, that they would be quickly brought before the judge for any violations, that they would receive a modest jail sanction for violations, and that they would not be able to bargain their way out of sanctions. They also understood that treatment and other assistance would be available as needed. ***Most cited the warning hearings, as well as briefings provided by their probation officers and other HOPE team members, as being crucial to***

[30] This finding may be due to the aspect of HOPE wherein the short jail sanctions are used in lieu of revocation.

imparting to them a solid understanding of HOPE. Thus, there seems to be little reason for concern that the HOPE probationers did not know what was expected from them in HOPE

Probationers' Violations Experiences

Exhibit 4-20 shows that 24% of HOPE probationers and 31% of PAU probationers reported at the Wave 1 interview that they had been found in violation. ***At the Wave 2 and Wave 3 interviews, the violation rate among HOPE probationers was significantly higher than PAU probationers.*** At Wave 2, about 63% of HOPE probationers reported that they had been found in violation, compared to about 30% of PAU probationers. At Wave 3, nearly 52% of HOPE probationers compared to about 30% of PAU probationers reported that they had been found in violation.

Exhibit 4-20. Self-reported past 6-month probation violations and incarceration by group and wave

	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
	HOPE	PAU	HOPE	PAU	HOPE	PAU
During the past 6 months, did the judge find you in violation of probation?	24.2	31.5	**63.2****	**29.6**	**51.9****	**29.8**
If you have met with your probation officer at least once since the start of your probation term, have you been required to spend time in jail in the past six months?	51.7	51.5	**70.0****	**35.3**	**62.7****	**37.1**
During the last 6 months, about how many days did you spend incarcerated in jail or prison? (mean)	45.1	42.2	**35.2****	**23.0**	**40.7***	**29.5**

* HOPE and PAU differ at $p < 0.05$; ** HOPE and PAU differ at $p < 0.01$

Probationers were also asked if they had been required to spend time in jail in the past 6 months, and how many days in the past 6 months they were incarcerated in jail or prison (***Exhibit 4-20***). The jail incarceration rate among probationers in both groups was similar at the Wave 1 interview, but at subsequent interviews, the jail incarceration rate among HOPE probationers was significantly higher than PAU probationers. At Wave 2, 70% of HOPE probationers, compared to about 35% of PAU probationers, reported that they had served time in jail in the previous 6 months. At Wave 3, 63% of HOPE probationers, compared to 37% of PAU probationers, reported that they had served time in jail in the previous 6 months. Similarly, at the Wave 1 interview, probationers in both groups reported spending about the same average number of days incarcerated in the past 6 months, but at subsequent interviews, HOPE probationers reported spending significantly more days incarcerated than PAU probationers. At Wave 2, HOPE probationers reported spending an average of about 35 days incarcerated in the past 6 months, compared to an average of 23 days for PAU probationers. At Wave 3, HOPE probationers reported spending an average of about 41 days incarcerated, compared to about 30 days on average for PAU probationers.

We asked the 21 HOPE interviewees if they had any violations during their time in HOPE and if so to describe them. Most respondents (N=18, or 86%) had experienced at least one violation, with the modal and mean number of violations being three. The maximum number of violations reported was six. As was to be expected given the heavy emphasis on drug testing within the HOPE DFE, the most common

reason for a violation was either a positive drug test, or the probationer missing a scheduled drug test. Tardiness for meetings with probation officers or to treatment groups was also commonly reported. One probationer reported being purposely late for a meeting to "test the limits of HOPE." Miscellaneous other reasons for violations included associating with prohibited persons (e.g., other felons), failure to maintain or to actively seek employment, and violating a PFA (which also resulted in an arrest). Many respondents who had multiple violations had different types of violations.

The probationers who violated universally reported that they received a sanction for the violation, most commonly a few days in jail. For those who reported multiple violations over time, they also reported that sanctions were progressive with each successive violation. Thus, we see evidence of the certainty principle being applied, with greater severity being reserved for repeat offenders. All of this is consistent with the underlying HOPE model.

Probationers' Attitudes and Thought Processes

Probationers' attitudes across six domains, including identification with criminals, tolerance for law violations, legal cynicism, readiness for change, self-efficacy, and substance abuse treatment motivation, were measured at each interview wave to assess change over time. ***Exhibit 4-21*** shows the average attitude scores in the six domains for both study groups at each wave. In general, each attitude score was derived by summing item-level responses at the individual level, then averaging by study group (see Appendix F for items and item-level responses for each domain).

Exhibit 4-21. Self-reported attitudes, by group and wave

Attitude Score	Wave 1 means		Wave 2 means		Wave 3 means	
	HOPE	PAU	HOPE	PAU	HOPE	PAU
CSS-M Identification with Criminal Others (1-12; higher score = greater identification with criminal others)	3.65	3.76	3.57	3.86	**3.61***	**4.11**
CSS-M Tolerance for law violations (0-20; higher score = greater tolerance)	4.52	4.90	**4.35****	**5.52**	**4.66***	**5.57**
Legal cynicism score (5-45, higher score=lower cynicism)	35.33	34.93	35.64	34.06	35.97	34.53
Readiness for change score (4-36; higher score=more ready for change	31.03	30.30	31.00	31.13	30.44	30.01
Readiness for change score, if incarcerated (5-45; higher score=more ready for change)	34.00	39.29	34.30	36.05	34.14	36.58
Self-efficacy score (4-36; higher score=greater self-efficacy)	27.99	27.52	28.39	28.46	**29.32****	**28.12**
Substance abuse treatment motivation score (10-90; higher score=more motivation)	41.91	42.14	39.50	42.06	**38.65****	**43.34**

** $p < 0.01$; * $p < 0.05$

"Identification with criminal others" was measured based on how strongly respondents agreed with six statements such as, "People who have broken the law have the same sorts of ideas about life as me." The identification score ranged from 0 to 12, where a higher score indicates a higher level of identification with crime-involved people. ***Exhibit 4-21*** shows that at each interview wave, HOPE

probationers expressed a lower level of identification with crime-involved people than PAU probationers, with a significant difference detected at the Wave 3 interview (3.61 versus 4.11).

To measure tolerance for law violations, probationers were asked to rate how strongly they agreed with 10 statements such as, "Sometimes a person like me has to break the law to get ahead in life." The tolerance score ranges from 0 to 20, where a higher score indicates greater tolerance for law violations. ***Exhibit 4-21*** shows that HOPE probationers expressed significantly lower tolerance for law violations than PAU probationers at the Wave 2 (4.35 versus 5.52, respectively) and Wave 3 interviews (4.66 versus 5.57, respectively).

Exhibit 4-21 also shows legal cynicism scores for both study groups at each interview wave. This score was derived from probationers' ratings about how strongly they agreed with four statements such as, "Laws are made to be broken." Scores range from 4 to 45 with a higher score indicating lower legal cynicism. Across all interview waves, scores were similar for the HOPE and PAU probationers.

Probationers who were in the community at the time of the interview also were asked to rate how strongly they agreed with four statements related to their readiness for change such as, "I am trying to stop committing crimes." The readiness for change score ranges from 4 to 36, where a higher score indicates greater readiness for change. ***Exhibit 4-21*** shows that there were no significant differences in readiness for change between HOPE and PAU probationers in the community. To assess readiness for change among those incarcerated, interview participants were asked to rate how strongly they agreed with five statements such as, "I want to get my life straightened out." This score ranges from 5 to 45, where a higher score indicates a greater readiness for change. Across interview waves, HOPE and PAU probationers had similar readiness for change scores (***Exhibit 4-21***).

Self-efficacy was measured by how strongly respondents agreed with four statements such as, "My life has gone out of control." The self-efficacy score ranges from 4 to 36, where a higher score indicates greater self-efficacy. ***Exhibit 4-21*** shows that at the Wave 1 and Wave 2 interviews, probationers in both study groups reported similar levels of self-efficacy. At the Wave 3 interview, HOPE probationers expressed a higher level of self-efficacy than PAU probationers (29.32 versus 26.12, respectively)—supporting HOPE team members' assessment that HOPE was helping participants manage their lives better.

Probationers were asked to rate how strongly they agreed with eight statements related to their motivation for substance abuse treatment such as, "You need help in dealing with your drug use." At the individual level, a treatment motivation score was derived by summing item-level ratings, dividing by the number of non-missing items, and multiplying by 10. The range for this score is 10 to 90, where a higher score indicates greater treatment motivation. At each interview wave, PAU probationers expressed a higher level of substance treatment motivation than HOPE probationers, with a significant difference detected at the Wave 3 interview (43.34 versus 38.65, respectively).

In summary, probationers in both study groups reported similar attitudes across the six domains at their baseline interview. ***In follow-up interviews, HOPE probationers expressed lower levels of identification with crime-involved people and tolerance for law violations.*** Legal cynicism was lower among HOPE probationers than PAU probationers but not significantly so. The level of readiness for change remained consistent for both study groups; however, ***HOPE probationers expressed a higher level of self-efficacy and a lower level of substance abuse treatment motivation than PAU probationers***.

For the 18 HOPE probationers interviewed who had at least one violation during HOPE, we asked if they thought about the possibility of being punished before they committed their violation, whether they had considered alternatives to their actions that led to the violation, and whether they felt that the sanctions they received would have any impact on their subsequent behavior. Our goal was to map the connection in their minds between thinking and action (Andrews and Bonta, 2003; Bucklen and Zajac, 2009), which would seem to be important to the deterrence message.

The responses by these 18 violators to the question of whether they thought about the consequences of their behavior beforehand fall into three broad categories:

- ***About half of the group reported that they had not thought much or at all about the consequences of their violations beforehand.*** Some of them argued that what they had done was not a violation anyway (or should not have been, such as being just a few minutes late for a meeting) or that they were railroaded by a corrupt system for something they did not do. Others reported that they were too dependent on drugs to rationally consider consequences (they could not resist using drugs, or were already high while committing a secondary violation). ***A hard core of this group indicated they not only did not think about the consequences, they would not have cared about them even if they had*** (or, alternately, they did think about consequences, but wanted to do what they wanted anyway). They were not troubled by the prospect of a few days in jail, believed they could "beat the system" (e.g. by tampering with their drug tests) and saw HOPE as illegitimate. One or two adamantly indicated that they "were not criminals" (despite high risk scores and documented criminal histories) and were "laying low" and biding their time on HOPE until they could be discharged and return to their favored activities (such as drug use). Several indicated they would move from their current county where HOPE was operating so that it would not again interfere with their lifestyles. Thus, for this first category, which we might dub the "***noncompliant***" we see some evidence of embedded patterns of anti-social attitudes, poor decision making and a general defiance towards authority. They did not seem to be impressed with the goals of HOPE and were generally not worried about the consequences HOPE might place on them. They did not feel HOPE helped them and often thought of HOPE as a "joke," as "corrupt," or as simply something to put up with until they could get out of it. Members of this category averaged two violations, although one member had six, and another three members had three violations.
- ***The second slightly smaller group consisted of probationers who initially did not think about the possibility of sanctions, but subsequently began to consider more carefully the impacts of their decision making on their lives.*** Several of these participants reported trying to "test" the system early on, for example by intentionally being late for appointments or by trying to "flush" their systems with water to fool the urinalysis drug tests. They indicated that over time they began to care more about going to jail and further complicating their lives and began to "get with the program." Some reported being tired of going to jail, weary of their criminal lifestyles, and ready for a change. Others were relatively new to crime and did not want to go any further into a criminal lifestyle. Thus, for this second category, which we might dub the ***"reformed,"*** we see evidence that the message of HOPE did get through to them. While initially resistant, they began to care about consequences and to try to

comply with program requirements. They spoke of the need to be accountable and to accept responsibility for their actions. Interestingly, this group had twice the mean number of self-reported violations (four) as did the "noncompliants," but without the higher end outliers. The "reformed" group reported that the violations they did have occurred early in the program, and that they matured into a period of violation-free behavior as they accepted the tenets of HOPE. They felt that the sanctioning regimen under HOPE helped them, but several also gave equal or even greater credit to intensive drug treatment programs or to other factors such as their religious beliefs.

- ***The third category consisted of only a few probationers who indicated that they thought a great deal about the consequences of violating, but did so anyway.*** These probationers indicated that their behavior was beyond their control, owing to factors such as extreme addiction. Given the very small number in this group, little more can be said about them.

Three of the probationers in our sample reported no violations during their time in HOPE (which we independently verified). For this group, the "***compliants,***" we asked if the possibility of punishment kept them from violating. Again, given that there were only three probationers in this group, we could gain only so much insight into their thought patterns. One indicated that fear of sanctions did have a big impact on decision making. Another reported that intensive drug treatment was more important than the prospect of sanctions. The third member of this group was somewhat vague on this question, suggesting that it was more about self-efficacy and will power than fear of punishment, although this person had clearly been through the criminal justice system over many decades and made it clear that returning to jail was no longer an option for them. ***Two of these three respondents seemed to be offenders who were aging out of crime and were ready to change. The third was a first-time offender who was clearly frightened of jail.*** Thus, HOPE seemed to be well received by members of this group.

One might speculate that the "noncompliants" were being helped by HOPE, given that they had fewer average violations than the "reformed" group. But, as noted earlier, several of them indicated that they were "going along to get along" and fully intended to resume their criminal activities after discharge from probation. They also indicated that if they got onto probation again, they would seek to get into PAU, which would allow them much greater freedom to do what they wanted. So, they either were being helped in some measure by HOPE or were clever enough to moderate their behavior temporarily.

Conversely, the "reformed" group reported that they believed HOPE was helping them, especially as they became more advanced in the program, but they had a relatively larger number of violations (albeit skewed towards their early days in HOPE). So, they may have been helped by HOPE, or they may have simply been engaging in socially desirable responding in their interviews, especially if they had become sympathetic to the goals of HOPE and had developed rapport with HOPE staff, which seemed to be the case.

Perceptions about Features of Probation and HOPE

At each ACASI interview wave, most HOPE and PAU probationers reported that regular face-to-face contact with their probation officer was a condition of probation, and that they had met with their probation officer at least once since the start of their probation term (***Exhibit 4-22***). At the Wave 2 and Wave 3 interviews, HOPE probationers reported a higher average number of meetings with their

probation officer than did PAU probationers. For example, HOPE probationers reported an average of 24 meetings with their probation officer, compared to about 14 meetings for PAU probationers.

Exhibit 4-22. Self-reported attitudes about probation officers, by group and wave

	Wave 1 means		Wave 2 means		Wave 3 means	
Attitudes about the Probation Officer	HOPE	PAU	HOPE	PAU	HOPE	PAU
Is regular face-to-face contact with your probation officer a condition of supervision?	93.8%	92.7%	**89.2%****	**95.7%**	88.6%	90.0%
Have you met with your probation officer at least once since the start of your probation term?	70.0%	65.3%	**93.4%***	**97.0%**	94.7%	93.7%
About how many times have you met with your PO?	7.6	7.0	**24.0****	**14.4**	**31.9****	**18.1**
Attitude about Probation Officer						
My probation officer is knowledgeable about my case.	6.43	6.54	6.77	6.94	**6.83***	**6.22**
My probation officer knows me by name.	**6.62***	**7.17**	**8.01***	**7.52**	**7.94****	**7.17**
My probation officer helps me to succeed.	5.85	6.20	5.98	6.24	6.08	6.12
My probation officer gives me a chance to tell my side of my story.	6.04	6.38	5.97	6.40	6.14	6.23
My probation officer treats me fairly.	6.39	6.62	**6.32****	**7.02**	6.45	6.65
My probation officer treats me with respect.	6.82	6.86	**6.65****	**7.32**	6.88	6.92
Score (7-54; higher score=positive attitude)	38.20	39.81	39.70	41.44	40.32	39.33

Note: Response ratings 1=strongly disagree to 9=strongly agree
** HOPE and PAU differ at $p < 0.01$; * HOPE and PAU differ at $p < 0.05$

Probationers were asked to indicate how strongly they agreed with six statements about their interactions with their probation officer (***Exhibit 4-22***). Ratings for these items ranged from 1 (strongly disagree) to 9 (strongly agree). Responses to these items were summed to create an overall score reflecting the probationer's attitude about their probation officer. This score ranges from 7 to 54, where a higher score indicates a more positive attitude toward the probation officer. At the Wave 1 and Wave 2 interviews, probationers in both study groups reported similar, relatively high levels of agreement that their probation officer is knowledgeable about their case. At the Wave 3 interview, HOPE probationers reported a higher level of agreement than PAU probationers that their probation officer is knowledgeable about their case (6.83 versus 6.22). At the Wave 1 interview, PAU probationers expressed a higher level of agreement than HOPE probationers that their probation officer knew them by name (7.17 versus 6.62, respectively). At the Wave 2 and Wave 3 interviews, HOPE probationers endorsed this statement with a higher level of agreement than PAU probationers. Probationers in both study groups and across interview waves gave generally neutral ratings to the statements, "My probation officer helps me to succeed" and "My probation officer gives me a chance to tell my side of the story." ***At the Wave 2 interview, HOPE probationers expressed lower levels of agreement than PAU probationers about being treated fairly by their probation officer and about being treated with***

respect by their probation officer. There was no significant difference between study groups with respect to these ratings at the Wave 3 interview. At each interview wave, the overall rating of probationers' interactions with their probation officer was similar for both study groups, with scores generally falling in the bottom end of the upper one-third of the score's range (39 to 54), indicating a modestly positive attitude toward the probation officers.

At each interview, probationers were asked about conditions of probation (e.g., attending treatment), drug testing experiences, and experiences with responses to their behavior while under supervision (e.g., reprimand by a judge). ***Exhibit 4-23*** shows that at Wave 1, probationers in both groups reported similar conditions of probation. For example, somewhat more than 60% probationers reported that keeping a job, going to school, or attending job training was a condition of supervision, while about 80% reported a drug-testing requirement. At Wave 2, conditions involving work/school, community service, and substance abuse treatment were also similar between study groups. However, a greater proportion of HOPE probationers compared to PAU probationers were required to take drug tests (89% versus 80%). (HOPE probation included random drug testing; most testing of PAU participants was scheduled or "for cause.") This finding held at the Wave 3 interview, where about 90% of HOPE probationers and 80% of PAU probationers reported that drug testing was a condition of supervision. Additionally, at Wave 3, a greater proportion of HOPE probationers compared to PAU probationers was required to keep a job, go to school, or attend job training (70% versus 57%).

Exhibit 4-23. Self-reported conditions of supervision, by group and interview wave

	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
Conditions of Supervision	**HOPE**	**PAU**	**HOPE**	**PAU**	**HOPE**	**PAU**
Is keeping a job, going to school, or attending job training a condition of supervision?	65.5	60.6	64.7	63.9	**69.5****	**57.1**
Is community service a condition of your supervision?	41.6	38.7	41.5	42.4	44.0	41.1
Is attending a drug or alcohol treatment program a condition of your supervision?	43.1	45.6	36.1	35.5	43.0	41.0
Is taking drug tests a condition of your supervision?	81.3	78.1	**89.0****	**79.6**	**89.5****	**80.1**

** HOPE and PAU differ at $p < 0.01$

Probationers with a drug testing requirement as a condition of supervision were asked more detailed questions about their drug testing experiences. ***Exhibit 4-24*** shows that at the Wave 1 interview, probationers in both groups reported the same average number of drug tests in the past 6 months (4.9). At the Wave 2 interview, HOPE probationers reported a significantly higher average number of tests than PAU probationers at Wave 2 (29.5 versus 8.2) and Wave 3 (20.8 versus 7.4). At Wave 1, probationers in both study groups reported similar positive drug test rates (44% of HOPE and 40% of PAU). At Wave 2, a significantly higher proportion of HOPE probationers than PAU probationers reported testing positive for drugs in the past 6 months (45% versus 29%), while there was no difference

in the positive drug test rate at Wave 3 (about 30% of each group). There was no difference between groups at any wave in the number of times tested positive (of those with at least one positive test).

Exhibit 4-24. Self-reported past 6-month drug testing experiences, by group and wave

	Wave 1		Wave 2		Wave 3	
Drug Testing Experience	HOPE	PAU	HOPE	HOPE	PAU	HOPE
If drug testing is a condition, how many times have you had a test in the past 6 months? (mean)	4.9	4.9	**29.5****	**8.2**	**20.8****	**7.4**
If drug tested, have you tested positive at least once in the past 6 months? (%)	44.0	39.8	**44.6****	**29.2**	29.4	30.8
If tested positive:						
How many times have you tested positive for drug use in the past 6 months? (mean)	2.2	3.0	2.2	4.1	4.8	2.4
Did you not receive a sanction (i.e., nothing happened)? (%)	22.0	14.1	3.6	6.5	7.9	13.5
Did you receive a verbal warning? (%)	21.1	27.3	**12.5****	**43.5**	**9.5****	**35.1**
Did you receive stricter or additional supervision conditions? (%)	20.2	17.2	21.4	30.4	20.6	35.1
Did you receive an increase in the frequency of AA/NA meetings? (%)	4.6	8.1	6.3	6.5	19.0	13.5
Did you receive an increase in the frequency of required drug or alcohol treatment? (%)	14.7	10.1	17.9	17.4	23.8	21.6
Did you receive an increase in drug treatment requirements? (%)	7.3	10.1	11.6	23.9	17.5	18.9
Did you receive a formal violation? (%)	17.4	14.1	20.5	21.7	25.4	24.3
Did you receive jail time as a sanction? (%)	36.7	47.5	**87.5****	**52.2**	**77.8****	**48.6**

** HOPE and PAU differ at $p < 0.01$.

Probationers who tested positive for drugs in the past 6 months were asked about the sanctions they received as a result. ***Exhibit 4-24*** shows that at Wave 1, both study groups reported receiving similar sanctions because of positive drug tests in the previous 6 months. For example, 21% of HOPE probationers and 27% of PAU probationers reported that they received a verbal warning from their probationer officer and 37% of HOPE and 48% of PAU probationers reported receiving jail time. At the Wave 2 and Wave 3 interviews, PAU probationers were much more likely to report receiving a verbal warning by the probation officer than were HOPE probationers compared to PAU probationers. In contrast, ***at Wave 2 and Wave 3, HOPE probationers were significantly more likely to report receiving jail time than the PAU probationers (88% and 52% at Wave 2; 78% and 49% at Wave 3). These results are consistent with the principles of HOPE.***

To better understand the experience of HOPE and PAU, probationers were asked whether they had experienced any of several common responses to their behavior while under supervision. ***Exhibit 4-25***

shows that at the Wave 1 interview, the past 6-month experiences for probationers in both study groups were similar, with the most common experience for both groups being jailed (about 52% of both groups). ***At Wave 2, past 6-month reported probation experiences were different for HOPE probationers compared to PAU probationers.*** HOPE probationers were significantly more likely to spend time in jail, be reprimanded or praised by a judge, receive a formal violation, be reprimanded by a probation officer, experience changes in the frequency of drug testing, and be placed on electronic monitoring or house arrest. At Wave 3, HOPE probationers continued to report a different probation experience than PAU probationers. Compared to PAU probationers, HOPE probationers were significantly more likely to spend time in jail, receive a formal violation, be praised by their probation officer, be reprimanded or praised by a judge, and experience changes in the frequency of drug testing. ***Most HOPE probationers compared to about one-third of PAU probationers reported jail time at both Wave 2 and 3.***

HOPE EXPERIENCES

HOPE probationers were much more likely than PAU probationers to report receiving jail time, interacting with a judge (praise or reprimand), receiving a formal violation, and experiencing a change in drug testing frequency during their first 12 months in the study.

Exhibit 4-25. Self-reported past 6-month probation experiences, by group and interview wave

	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
Probation Experiences	HOPE	PAU	HOPE	PAU	PAU	HOPE
Placed on house arrest in the past 6 months	11.4	11.5	**11.0***	**5.8**	10.7	6.7
Required to do community service	43.3	45.2	45.2	44.6	38.1	36.0
PO increased frequency of drug tests	21.4	15.6	**23.6****	**11.4**	**31.7****	**14.0**
PO decreased frequency of drug tests	8.6	6.4	**40.9****	**19.5**	**46.2****	**23.2**
PO increased frequency of SA treatment	15.7	15.9	17.6	12.2	20.2	14.0
PO decreased frequency of SA treatment	6.2	4.8	13.2	9.5	16.7	15.7
Reprimanded or warned by probation officer	**34.7***	**42.6**	**52.1****	**39.7**	46.6	46.1
Praised by probationer officer	24.2	28.2	53.6	45.9	**57.1***	**44.9**
Reprimanded or warned by judge	32.7	30.5	**60.6****	**16.7**	**49.8****	**27.7**
Praised by judge	11.3	9.7	**25.6****	**10.0**	**33.9****	**14.2**
Received formal violation	33.9	39.5	**57.5****	**35.3**	**49.0***	**37.1**
Required to spend time in jail	51.7	51.5	**70.0****	**35.3**	**62.7****	**37.1**

** HOPE and PAU differ at $p < 0.01$; * HOPE and PAU differ at $p < 0.05$

We asked the 21 HOPE respondents we interviewed which aspects of HOPE they felt were the most helpful to them, and what they felt HOPE did best. ***Approximately two-thirds of the respondents***

indicated that the swift and certain sanctioning of HOPE was helpful. They emphasized themes of strictness, accountability and the expectation of personal responsibility as key features of HOPE and saw this as quite distinct from the leniency and inconsistency which they felt was characteristic of PAU. HOPE provided to them more certainty, predictability and clearer expectations than PAU. ***Several of them also explicitly noted the inconvenience of HOPE – that if they commit a violation their lives would be disrupted by a short jail stay – as being an important motivator in changing their behavior.*** This is line with the espoused theory of HOPE, where sanctioning is not intended to be severe, but simply sufficiently disruptive to the normal routines of probationers to get their attention (Hawken and Kleiman, 2009). ***The desire to avoid jail was also frequently mentioned, either because they were afraid of it or because they were tired of it.*** Still, over half of these respondents also indicated that while the swift and certain sanctioning routine of HOPE was valuable, they felt that other factors, such as treatment programs or even their spiritual lives, were equally or even more important to their rehabilitation. This was especially so ***for those who reported being serious addicts. These respondents felt strongly that sanctioning by itself did not matter, that intensive treatment was what helped them the most.*** Thus***, while we see endorsement of the swift and certain sanctioning approach under HOPE, many of the respondents also felt it was not enough, or not even the primary factor, in their attempts at going straight.*** These observations about the strengths of HOPE were also widely echoed in our interviews with the HOPE team members, suggesting a common understanding of HOPE among those who were running HOPE and those who were participating in it.

This group of participants was optimistic that HOPE would help them stay out of trouble in the future, and that HOPE more broadly had a positive impact on their ability to manage their lives more generally. They felt that the strict demands of HOPE had improved their skills with time management, financial responsibility, accountability to others (e.g., employers and family), interpersonal communication and other life skills. HOPE had provided them with a degree of structure that heretofore had been missing in their lives. Some felt that HOPE had helped them to "grow up." But, others were apprehensive about what would become of them once this externally imposed structure was removed. Some were requesting that they be allowed to remain on the drug testing color line even after they had been stepped down off it, and others wished to remain on it even after they completed probation. They feared that they would relapse to substance use or other criminal behavior without the close oversight provided by HOPE. Similar findings were reported in the Hawaii HOPE study (Hawken and Kleiman, 2009). This raises the question of the extent to which HOPE is building resilience in its clients, allowing them to maintain their gains even after discharge from the program. We also heard similar themes and concerns in interviews with HOPE team members, indicating agreement between program providers and recipients.

Another third of respondents felt that there was little of value to HOPE. This group evidenced stronger levels of criminal thinking, more oppositional defiance and a general reluctance to accept oversight by probation. For example, they stated

- They did not care about jail sanctions (and that they knew other HOPE participants who were also indifferent towards jail);
- They had complaints about various aspects of HOPE, such as their relationships with their probation officers and the judge;

- The drug testing schedule was excessive; and
- HOPE in general was too inflexible.

Among this group, some expressed a preference for PAU, where they believed they could do as they wished (thereby validating the impression of PAU garnered from the HOPE teams that PAU was quite lax compared with HOPE). While some of them felt that treatment programs were of more help to them, not all were sanguine about treatment either.

We also asked the respondents what parts of HOPE were the most difficult. By far the most frequently mentioned was the drug testing color line and the frequency of drug testing. HOPE probationers were required to call in early to the color line to determine if they were required to report for testing that day; and, if they were, in some cases they needed to report for testing by noon. Many respondents reported that this posed a problem for them. For some, the reasons were frivolous, such as preferring to sleep in until the early afternoon. But for many others, the conflict was related to employment. This was most notable for construction workers who had to report for work as early as 6am and did not clock out until 2 or 3pm. Thus, they were working during the entire time when the drug testing centers at the probation offices were open. In addition, these workers were often at distant job sites, which could change from week to week, and they often lacked reliable transportation. Some employers were accommodating and would allow a probationer to report late for work or to quit early to comply with testing, but others were not. ***Several respondents reported having lost jobs due to the random drug testing, sometimes more than once. For homeless probationers, maintaining access to a telephone to be able to call into the color line was also a challenge.*** The HOPE teams were aware of this problem and mentioned it often during our interviews with them. The teams did appear to want to make accommodations where needed, such as extending testing hours and even engaging in outreach with employers to encourage them to be flexible with these probationers. This outreach was sometimes effective and sometimes not. Thus, ***the focus on intensive surveillance in HOPE often ran up against the equally important goal of maintaining steady employment*** for this population.

PROBLEMS POSED BY HOPE REQUIREMENTS

The 21 HOPE participants who discussed HOPE with the evaluation team identified the frequent, random drug testing as the most difficult component of HOPE because the testing requirements interfered with their employment.

Another theme mentioned by the respondents was ***the difficulty of giving up drugs, changing their way of thinking, and distancing themselves from old friends who would counter the accountability message they were receiving from HOPE.*** But, the probationers viewed this as part of the challenge and often noted that their probation officers and other members of the HOPE team tried to help them work through on these issues. They also often ***credited drug treatment programs in which they participated during HOPE as being critical to cognitive transformation***.

Fairness of Sanctions

As noted above, nearly 25% of HOPE probationers and about 32% of PAU probationers reported at the Wave 1 interview that they had been found in violation. ***Exhibit 4-26*** shows that about one-third of probation violators in each group reported that the violations were a surprise (35% of HOPE and 32% of

PAU), and nearly 40% in each group thought the violations were unfair (38% of HOPE and 39% of PAU). At Wave 2, about 63% of HOPE probationers reported that they had been found in violation, compared to about 30% of PAU probationers. Despite the higher violation rate among HOPE probationers compared to PAU probationers, there was little difference between violators in the study groups in reporting that the violation was a surprise and similar proportions of both groups said the violation was unfair. At Wave 3, nearly 52% of HOPE probationers compared to about 30% of PAU probationers reported that they had been found in violation, but like the Wave 2 findings there was no significant difference in the proportion who felt the violation was a surprise or that the punishment was unfair.

Exhibit 4-26. Perceived fairness of sanctions among interview participants, by group and wave

Fairness Items	Wave 1 (%)		Wave 2 (%)		Wave 3 (%)	
	HOPE	PAU	HOPE	HOPE	PAU	HOPE
If you were found in violation of probation, did the violation that you received come as a surprise to you?	35.1	32.4	32.8	32.4	32.6	24.6
If you were found in violation of probation, do you think the punishment you received for violating probation was unfair?	37.7	39.3	44.1	37.9	40.6	42.9

There were no differences between HOPE and PAU.

During the qualitative interviews, we asked the respondents whether they thought that HOPE was fair. Fairness is the important third leg of the swift-certain-fair description of HOPE principles. Fairness is related to the idea that sanctions should not be overly severe, but also to probationers' more global estimation of how they feel they have been treated in HOPE. Like findings from the ACASI interview , ***most of the HOPE interviewees felt that HOPE was fair.*** Some drew distinctions among the various parts of HOPE, for example, indicating that the judge was fair, but not the probation officers, or vice versa. Several indicated that success in HOPE is entirely up to you, and that if you think it is unfair it is because you are not trying hard enough to succeed. Others felt that while HOPE was fair, it was sometimes too intolerant of innocent errors or factors beyond the control of the individual probationer, such as being late for an appointment because of a traffic tie up. Several respondents reported that they knew of some HOPE probationers who had "gotten away" with things that others were punished for, thus raising some concerns about consistency. ***Several respondents indicated that they felt HOPE was unfair, that the program and the HOPE team was "out to get them" and that HOPE "sets you up to fail."*** But, these sentiments were the distinct minority. Few felt that HOPE was doing them any real or subjective harm. Thus, the HOPE DFE seems to have been successful in cultivating a sense of fairness among its participants.

4.7. Plans for Sustainability

Given the positive attitudes towards HOPE and the overall ease of implementation reported at most sites, it is not surprising that most members of the HOPE teams voiced a desire to see HOPE continue at their sites. As of the final site visits in the fall of 2014, though, none of the sites had developed detailed plans for continuation. Tarrant County operated the SWIFT program prior to HOPE, and the intention there was to merge the two after the DFE and press on under the rubric of SWIFT. Essex County and

Saline County had begun to explore options for building legislative support in their states for HOPE to leverage continued funding. But again, no definitive plans had been finalized. To their credit, the sites were also very keen to see the results of the DFE before making firm plans for the longer-term future of HOPE, first wanting to know whether HOPE works.

Again, the one exception is Clackamas County. As discussed, there was conflict between HOPE and the pre-existing probation framework. By the time of the last site visit, attitudes towards HOPE had warmed considerably, and most team members reported seeing the value (at least in principle) of HOPE. Still, there was wide sentiment that while HOPE may have general merit, it was still not a good "fit" for their jurisdiction.

4.8. Summary of Process Evaluation Findings

The process evaluation measured implementation fidelity using data gathered by the local HOPE project coordinators hired by the sites and supervised by the TTA provider. Process data were collected during three rounds of site visits to the four sites that included interviews with HOPE stakeholders and, during the final round of visits, HOPE probationers. Attitudes towards HOPE and PAU were also assessed through questions included in the three waves of ACASI interviews with evaluation participants.

Findings showed that ***the four DFE sites implemented HOPE with fidelity*** based on measurement against eleven key metrics central to the underlying HOPE model (e.g. swift, certain, fair). All sites met at least a 60% standard on all metrics except for holding a violation hearing within 3 days of the violation, which was a struggle for all sites (only one site, Texas, achieved the 60% standard on this metric). There was agreement across the sites that this standard was difficult to achieve because of large geographic areas to cover, lack of adequate resources to service warrants, or the challenge in locating absconders. Although the sites missed the 3-day mark, further examination of the data showed that more than three-quarters of all violators did have a hearing within one week (7 days) of the violation.

Implementation fidelity was promoted by several factors:

- A training and technical assistance (TTA) provider was contracted by BJA to provide regular support and guidance to the sites. This provider, Pepperdine University led by Dr. Angela Hawken, was a regular presence at the sites and provided ongoing feedback and correction. Moreover, the provider also sometimes included Judge Alm in these visits.
- Staff from BJA also monitored program implementation.
- Each site had a full-time HOPE project coordinator, paid through the site's BJA grant award, whose role was to assure adherence to the HOPE principles; these individuals also collected the fidelity data for the TTA provider on a real-time basis.

Evaluation team interviews with HOPE stakeholders during three rounds of site visits revealed that, overall, ***there was strong buy-in to the HOPE concept.*** Those implementing the program believed in the model and were optimistic that HOPE would be successful. ***Implementation was facilitated by existing agency cooperation, prior experience with HOPE-like programs, and organizational linkages between probation and the court.*** The local administrative structure of probation was also important to implementation. In three sites, the probation department was either directly under the control of the HOPE judge or there was a sufficiently close administrative linkage such that the judge could substantially direct the operations of probation in the service of the HOPE program—this connection facilitated the establishment of HOPE and its operation. In the fourth site, probation was operated

locally from the Sheriff's Department and the HOPE judge had no administrative connection to probation which resulted in challenges. Other challenges in some sites included resource constraints—even with grant funding—and conflicts with existing probation culture (e.g. risk-needs-response versus the surveillance required by HOPE).

HOPE providers stated that implementing and operating HOPE was not overly burdensome and that any burden was "worth it." Stakeholders also consistently reported that they believed that HOPE was helping probationers to learn how to better manage their lives more generally through the setting and enforcement of expectations. ***Probationers, who had positive views of HOPE, also stated that they believed that HOPE had helped them better manage their lives***. HOPE team members and probationers noted that the frequent drug testing regimen could create conflicts for probationers who were employed and who lived (or worked) distant from the testing site. In some cases, probationers lost jobs because of the testing requirements.

HOPE team members also expressed some concerns that HOPE did not work as well for lower functioning or mentally ill probationers and for the more seriously antisocial probationers who were not threatened by a few days in jail . This latter view was confirmed by some of the interviewed probationers who reported that they were just going along with the program until their probation ended at which time they would go back to their antisocial lifestyles.

HOPE probationers understood what was expected of them. Both study groups had a strong sense that their probation officer would find out about noncompliance and would arrest them or have them arrested for noncompliance. Both groups also had a strong sense that the judge would do something in response to noncompliance, although HOPE probationers at their 12-month interview were more certain than PAU probationers that the judge would respond suggesting that the HOPE probationers—if they didn't understand initially—learned that sanctions would happen.

HOPE probationers who participated in a qualitative interview were mixed as to whether they thought about the potential consequences before committing violations. Most did <u>not</u> report giving the possibility of punishment much thought (even though they knew in an intellectual sense what <u>could</u> happen), with some suggesting that they did not actually care much about being punished. For some, though, the deterrence message set in over time, leading them to be more thoughtful about their behavior. Responses to ACASI interviews with probationers underscore this point: ***HOPE probationers were more sensitive to the possible consequences of noncompliance (as measured by the deterrence score) and reported a lower tolerance for law violations than their PAU counterparts.*** The ACASI interviews also offer some evidence about change in other attitudes among HOPE probationers. Specifically, at follow-up, ***HOPE probationers reported greater self-efficacy and a lower level of identification with crime-involved people than PAU probationers***.

ACASI interviews show that ***HOPE and PAU probationers experienced probation differently.*** Although HOPE and PAU probationers were equally likely to be required to attend substance abuse treatment as a condition of supervision, HOPE probationers were more likely to attend treatment. More HOPE probationers than PAU probationers were subjected to drug testing as a supervision requirement and very few PAU probationers were subjected to random testing. HOPE probationers who participated in a qualitative interview felt that ***the most difficult part of HOPE was balancing the need to report for frequent drug tests with their work schedules***, leading some to lose jobs due to their participation in

HOPE. ***They also felt that HOPE's emphasis on accountability was helpful, as was the structure it provided,*** which was often lacking before HOPE.

A final consideration is that some HOPE probationers came to rely on the drug testing hotline to maintain their sobriety and were afraid that once they were no longer tested that they would no longer be able to maintain sobriety. A similar finding was also reported by Hawken and Kleiman (2009).

5. Findings from the Outcome Evaluation

The purpose of the outcome evaluation was to determine whether HOPE improved outcomes for supervised populations compared to individuals who were on PAU or probation as usual. We examined the following outcomes by comparing HOPE participants to PAU participants using administrative data (i.e., including all study participants):

- Does HOPE participation reduce recidivism, measured by arrest, conviction, and probation revocation?
- Does HOPE participation improve compliance with conditions of supervision and reduce violations?
- What is the impact of HOPE on jail days served and prison days sentenced?

We examined the following using interview data:

- What is the impact of HOPE on drug use?
- Does HOPE participation change potential mediators including dynamic recidivism risk factors such as employment and housing stability?
- Does HOPE participation change attitudes that are potential mediators, including participants' criminal thinking/attitudes, perceptions of locus of control, and perceptions of the criminal justice system fairness/legitimacy?

We also examined the relationship between violations and sanctions for the HOPE participants only, using administrative data.

5.1. Impact of HOPE on Recidivism

Some of these findings were originally reported in Lattimore, et al. (2016).

Recidivism outcomes were new arrest (or arrest charge), probation revocation, revocation or arrest, and new conviction. The average length of follow-up was 650 days or more than 21 months, measured as the number of days from study intake to the end of our follow-up period (March 31, 2015); follow-up ranged from 186 to 969 days. We look first at the incidence of each of these recidivism outcomes and then results from survival models of time to first event and negative binomial count models of numbers of events.

Recidivism: Arrest

Results for new arrests over all sites are shown in ***Exhibit 5.1***. Overall, 42% of the study subjects were arrested at least once and there was no across-site variation ($F = 1.40$; $p = 0.241$). There were significant across-site differences in the likelihood of offense-specific arrest charges, with significant differences across-site in the likelihood that study participants received a person charge ($F = 8.89$; $p < 0.0001$), a property charge ($F = 3.56$; $p = 0.0137$), a drug charge ($F = 5.70$; $p = 0.0007$), or a public order/other charge ($F = 4.83$; $p = 0.0024$).[31] There was also variation across the sites in the average

[31] Overall site-level prevalence for any person charge was 0.13 (sd = .34) for AR, 0.16 (sd = 0.36) for Massachusetts, 0.08 (sd = 0.27) for OR, and 0.06 (sd = 0.23) for TX; for any property charge were 0.20 (sd = 0.40) for AR, 0.16 (sd – 0.37) for Massachusetts, 0.22 (sd= 0.410 for OR, and 0.14 (sd = 0.34) for TX; for any drug charge were 0.11 (sd = 0.32) for AR, 0.10 (sd = 0.30) for Massachusetts, 0.19 (sd = 0.40) for OR, and 0.14 (sd = 0.34) for TX;

numbers of recidivism arrests (F = 4.53; p = 0.0036)—which ranged from a low in Texas of 0.58 (sd = 0.90) to 0.75 (sd = 1.14) in Massachusetts to 0.83 (sd = 1.37) in Arkansas to 0.90 (sd = 1.47) in Oregon.

Exhibit 5-1. Recidivism arrest outcomes of HOPE DFE evaluation participants (mean and standard deviation)

Characteristic	Overall	HOPE	PAU	*t* Statistic
Length of follow-up (days)* **	650.0 (212.0)	653.8 (211.6)	646.4 (212.4)	–0.680
Recid any arrest charge	0.42 (0.49)	0.40 (0.49)	0.44 (0.50)	1.66
Recid person charge* **	0.11 (0.31)	0.10 (0.30)	0.11 (0.32)	1.070
Recid property charge*‡	0.18 (0.38)	**0.15‡ (0.36)**	**0.20 (0.40)**	2.647
Recid drug charge*†**	0.14 (0.34)	**0.12† (0.32)**	**0.15 (0.36)**	1.969
Recid public order/other**	0.28 (0.45)	0.27 (0.44)	0.28 (0.45)	0.403
Number recidivism arrests**	0.76 (1.24)	0.70 (1.22)	0.82 (1.26)	1.920
N	1,496	737	759	—

**Sites differ on measure at $p < 0.01$; ***Sites differ on measure at $p < 0.001$.
‡HOPE and PAU differ at $p < 0.01$.
†HOPE and PAU differ at $p < 0.05$.

Overall, there were no differences between the HOPE and PAU groups in the likelihood of new arrest charges during follow-up (t = 1.66; p = 0.11). More than 40% of both groups (42%) experienced at least one arrest during the follow-up period. ***The HOPE group was less likely than the PAU group to have been arrested for a property offense (15% vs. 20%) or a drug charge (12% vs. 15%). The HOPE group was equally likely to have experienced an arrest with a person charge or a public order/other charge at arrest.*** As there were not significant differences in the number of recidivism arrests or the likelihood of an arrest, the significantly fewer property and drug arrests for HOPE probationers appears to be a result of HOPE cases having slightly fewer arrests (0.7 vs. 0.82 on average) and being slightly less likely to be arrested (41% vs. 44%) although these differences were not statistically significant.

There were no differences overall between HOPE and PAU groups in the number of arrest charges (average 0.76), but there were within-site differences that favored the HOPE groups in two sites (Exhibit 5.2). Specifically, in Arkansas and Texas, the HOPE probationers experienced fewer recidivism arrests (68% versus 98% in Arkansas and 48% versus 68% in Texas). The likelihood of having at least one new arrest was similar across sites and groups—about 42%.

and for any public order/other charge were 0.30 (sd = 0.46) for AR, 0.29 (sd = 0.46) for Massachusetts 0.30 (sd = 0.46) for OR, and 0.20 (sd = 0.40) for TX.

There were across-site differences in arrest by offense type, with significant variation for all four offense types. The prevalence of a person charge ranged from about 5% (Texas PAU) to 16% (Massachusetts PAU); the prevalence of a recidivism property charge ranged from about 10% (Texas HOPE) to 25% (Oregon PAU); the prevalence of a recidivism drug charge ranged from about 8% (Arkansas HOPE and Massachusetts HOPE) to about 21% (Oregon HOPE); and the prevalence of a public order/other recidivism charge ranged from 19% (Texas PAU) to 32% (Arkansas PAU). ***The only significant within-site difference between groups on offense-specific arrest charge was observed in Texas, where HOPE probationers were significantly less likely to have experienced a recidivism drug charge (9% versus 17%).*** There was also across-site variation in the numbers of new arrests with the average number of new arrests lower in Texas than in the other sites (F = 4.53; p = 0.004).

Exhibit 5-2. Recidivism arrest outcomes by site and group (mean and standard deviation)

Site	Group	Length of follow up (days)***	Any Charge	Person Charge***	Property Charge*	Drug Charge***	Public Order/ Other**	Number Arrests**	N
AR	HOPE	619.4 (245.7)	0.38 (0.49)	0.11 (0.32)	0.17 (0.38)	0.08 (0.28)	0.28 (0.45)	**0.67† (1.14)**	178
	PAU	592.4 (244.2)	0.44 (0.50)	0.15 (0.36)	0.23 (0.42)	0.14 (0.35)	0.33 (0.47)	**0.99 (1.57)**	162
MA	HOPE	577.5 (199.6)	0.43 (0.50)	0.15 (0.36)	0.15 (0.36)	0.09 (0.28)	0.30 (0.46)	0.82 (1.32)	188
	PAU	590.8 (199.0)	0.48 (0.50)	0.16 (0.37)	0.17 (0.37)	0.11 (0.32)	0.29 (0.45)	0.69 (0.94)	203
OR	HOPE	723.9 (136.5)	0.43 (0.50)	0.06 (0.23)	0.18 (0.38)	0.21 (0.41)	0.29 (0.46)	0.83 (1.50)	190
	PAU	723.6 (139.5)	0.45 (0.50)	0.09 (0.29)	0.25 (0.44)	0.18 (0.39)	0.31 (0.47)	0.96 (1.44)	204
TX	HOPE	693.4 (220.9)	0.37 (0.48)	0.06 (0.24)	0.11 (0.31)	**0.09‡ (0.29)**	0.20 (0.40)	**0.48† (0.76)**	181
	PAU	668.8 (232.6)	0.41 (0.49)	0.05 (0.22)	0.17 (0.38)	**0.17 (0.38)**	0.19 (0.40)	**0.68 (1.00)**	190

**Sites differ on measure at $p < 0.01$; ***Sites differ on measure at $p < 0.001$.
‡HOPE and PAU differ at $p < 0.01$.
†HOPE and PAU differ at $p < 0.05$

Recidivism: Revocation

Hope supervision is theorized to lead to a reduction in probation revocations. Such a finding was not observed in the DFE. ***Overall, 24% of all DFE participants experienced at least one revocation and there was no difference between those on HOPE (26%) and those on PAU (22%; t = -1.837, p = 0.07). Exhibit 5.3*** shows that there was variation in revocation across the sites. Overall, the revocation rates (HOPE and PAU combined) were 23%, 21%, 13%, and 38% (Arkansas, Massachusetts, Oregon, and Texas; F = 24.32, p < 0.0001). There were also significant differences in revocation rates between HOPE and PAU groups in two sites. ***HOPE probationers were significantly more likely to be revoked in Arkansas and Oregon than those in the PAU groups.*** It should be noted that the PAU revocation rates were very low

in these two sites—about 13% in Arkansas and 9% in Oregon—suggesting that there was little 'room' for HOPE to have significantly lowered revocations in these two sites. Revocation was most likely in Texas—where 41% of the PAU and 35% of the HOPE probationers were revoked. HOPE probationers were somewhat less likely to be revoked in Massachusetts (19% vs. 23%; t = 0.88) and Texas (35% vs. 41%; t = 1.35), although these differences were not significant.

Exhibit 5-3. Revocation and revocation or arrest outcomes by site and group (mean and standard deviation)

Site	Group	Has probation revocation***	Has revocation or arrest	N
AR	HOPE	**0.33‡ (0.47)**	0.49 (0.50)	179
	PAU	**0.13 (0.34)**	0.45 (0.50)	163
MA	HOPE	0.19 (0.39)	0.44 (0.50)	189
	PAU	0.23 (0.42)	0.50 (0.50)	203
OR	HOPE	**0.17† (0.38)**	0.48 (0.50)	190
	PAU	**0.09 (0.29)**	0.47 (0.50)	204
TX	HOPE	0.35 (0.48)	0.52 (0.50)	185
	PAU	0.41 (0.49)	0.55 (0.50)	191

***Sites differ on measure at $p < 0.001$.
†HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.
Note: There is a small difference between the numbers of revocations and arrests for Arkansas HOPE and PAU reported here and reported in Lattimore, et al. (2016). In Lattimore, et al., HOPE was reported as 0.50 and PAU as 0.44—in contrast to the above values of 0.49 and 0.45. This small difference was due to an adjustment to the study intake date for a small number of cases.

Recidivism: Arrest or Revocation

Because revocation can foreclose the opportunity for a new arrest (as individuals are removed from the street) and are sometimes in lieu of arrest (a probation officer determines that the individual has committed a new crime and rather than arrest proceeds with revocation), it is useful to look at this combined measure of arrest/noncompliance with supervision conditions. ***Overall, the HOPE and PAU groups were similar on this measure—0.49 (sd = 0.50) of the HOPE group and 0.50 (sd = 0.50) of the PAU group experienced either revocation or arrest*** (or both; t = 0.370, p = 0.71).

There was also no significant variation in the combined arrest/revocation measure across the sites (***Exhibit 5-3***; F = 1.63, p = 0.18). Site-level rates (both groups combined) were 47% in Arkansas and Massachusetts, 48% in Oregon, and 54% in Texas. ***There were no differences between HOPE and PAU groups in the combined measure in any of the four sites.***

Recidivism: New Conviction

Exhibit 5-4 shows the percentages of DFE participants who were convicted of a new offense, overall and by HOPE and PAU status. ***There were no differences between HOPE and PAU participants in the proportions who had new conviction for any type of offense or in the proportions having a new conviction for person, property, drug, or public order/other offenses***. Overall 27% of the DFE participants had at least one new conviction. About 5% had a new conviction with a person charge, 11% with a property charge, 8% with a drug charge, and 13% with a public order/other charge. (Offense-specific convictions don't total to any conviction because some individuals had more than one offense associated with a conviction.) ***There was also no difference between groups in the average number of new convictions***—overall the DFE participants had an average of 0.38 new convictions.

Exhibit 5-4. New convictions overall and by group (mean and standard deviation)

Recidivism Event	Overall	HOPE	PAU	*t* Statistic
Conviction* **	0.27 (0.44)	0.28 (0.45)	0.26 (0.44)	–0.780
Person conviction	0.05 (0.22)	0.05 (0.23)	0.05 (0.21)	–0.710
Property conviction* **	0.11 (0.31)	0.11 (0.31)	0.11 (0.31)	0.066
Drug conviction* **	0.08 (0.28)	0.08 (0.27)	0.09 (0.29)	0.971
Public order/other conviction* **	0.12 (0.33)	0.13 (0.33)	0.12 (0.33)	–0.194
Number of convictions* **	0.38 (0.71)	0.38 (0.71)	0.37 (0.72)	–0.382
N	1,496	738	758	—

***Sites differ on measure at $p < 0.001$.
Note: There were no statistically significant differences ($p < 0.05$) between HOPE and PAU groups.

The sites differed in the proportion of DFE participants who had any conviction (***Exhibit 5-5;*** F = 14.35, p < 0.0001). Texas had the lowest rate of new convictions with 0.18 (sd = 0.38), followed by Massachusetts (0.23, sd = 0.42), Arkansas (0.31, sd = 0.46), and Oregon (0.42, sd = 0.37). ***The sites were similar in relatively low rates of new convictions for person offenses*** (Arkansas, 0.07, sd = 0.25; Massachusetts, 0.06, sd = 0.24; Oregon, 0.04, sd = 0.20; Texas, 0.03, sd = 0.18; F = 1.81, p = 0.14). ***Differences were observed across the sites for the rates of convictions for other types of offenses.*** Texas (0.06, sd = 0.24) and Massachusetts (0.07, sd = 0.26) had the lowest rates of convictions for property offenses, with Arkansas (0.14, sd = 0.35) and Oregon (0.17, sd = 0.37) somewhat higher (F = 11.18, p < 0.0001). Convictions for drug offenses were also lowest in Massachusetts (0.04, sd = 0.20) and Texas (0.05, sd = 0.23), with Arkansas (0.10, sd = 0.30) and Oregon (0.14, sd = 0.34) higher (F = 9.21; p < 0.0001). Conviction rates for public order/other offenses were lowest in Texas (0.05, sd = 0.23) and Arkansas (0.09, sd = 0.28), with Massachusetts (0.14, sd = 0.35) and Oregon (0.21, sd = 0.40) higher (F = 16.00, p < 0.0001). Finally***, there were significant across-site differences in the average number of new convictions*** (F = 21.88, p < 0.0001). The average numbers of conviction events were 0.21 (sd = 0.48) in Texas, 0.30 (sd= 0.60) in Massachusetts, 0.41 (sd = 0.69) in Arkansas, and 0.60 (sd = 0.59) in Oregon

Exhibit 5-5 also shows that there were three significant site-level differences between the HOPE and PAU groups with respect to new convictions. ***HOPE participants in Arkansas were more likely to have a recidivism conviction (t = -3.44) and to have more convictions on average (t = -2.61) than PAU***

participants. HOPE participants were less likely than PAU participants to have a new conviction for a drug offense in Texas (t = 2.22).

Exhibit 5-5. New conviction outcomes by site and group (mean and standard deviation)

Site	Group	Any***	Person	Property***	Drug***	Public Order/ Other***	Count ***	N
AR	HOPE	**0.39‡ (0.49)**	0.08 (0.28)	0.17 (0.38)	0.11 (0.32)	0.11 (0.31)	**0.50‡ (0.70)**	179
	PAU	**0.22 (0.41)**	0.04 (0.20)	0.11 (0.32)	0.09 (0.28)	0.06 (0.24)	**0.30 (0.66)**	161
MA	HOPE	0.23 (0.42)	0.06 (0.25)	0.06 (0.25)	0.04 (0.19)	0.14 (0.35)	0.30 (0.63)	188
	PAU	0.24 (0.43)	0.06 (0.24)	0.08 (0.27)	0.05 (0.22)	0.14 (0.35)	0.29 (0.58)	203
OR	HOPE	0.35 (0.48)	0.04 (0.20)	0.14 (0.35)	0.13 (0.33)	0.19 (0.40)	0.55 (0.90)	190
	PAU	0.39 (0.49)	0.04 (0.21)	0.19 (0.39)	0.14 (0.35)	0.22 (0.41)	0.63 (0.95)	204
TX	HOPE	0.16 (0.37)	0.03 (0.16)	0.06 (0.24)	**0.03† (0.16)**	0.06 (0.23)	0.18 (0.45)	181
	PAU	0.19 (0.39)	0.04 (0.19)	0.06 (0.23)	**0.08 (0.27)**	0.05 (0.22)	0.23 (0.50)	190

***Sites differ on measure at $p < 0.001$. †HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.

Recidivism: Time to First Arrest, Revocation, Arrest/Revocation, and Conviction

The bivariate analyses discussed above do not accommodate the fact that individuals were "on the street" for variable lengths of time or that different individuals had different risk exposures for the recidivism events depending on other events—for example, an individual in long-term residential drug treatment would, presumably, be at lower risk of committing new crimes or being revoked or arrested that someone in the community. Survival (or hazard) models allow us to accommodate these factors.

Exhibit 5-6 shows the Kaplan-Meier survival curves for the HOPE and Pau groups (overall) for time to first new arrest (charge). As the red HOPE line sits largely above the black PAU line, this suggests longer times to rearrest for HOPE probationers compared with PAU; however, the difference is not significant.

Exhibit 5-6. Kaplan Meier survival curves for time to first new arrest (charge) for all HOPE and PAU probationers

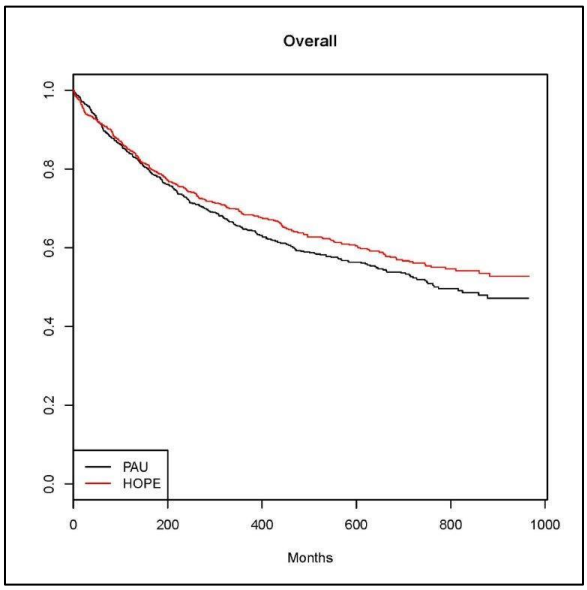


Exhibit 5-7 shows the survival curves for time to first new arrest (charge) for the HOPE and PAU groups by site. As can be seen, there are across-site differences, with the time to first new arrest quite similar for the two groups in Massachusetts and Oregon and somewhat longer times to first new arrest for the HOPE groups in Arkansas and Texas, although again these differences are not significant.

Exhibit 5-7. Kaplan-Meier survival curves for time to first new arrest (charge) by site and group

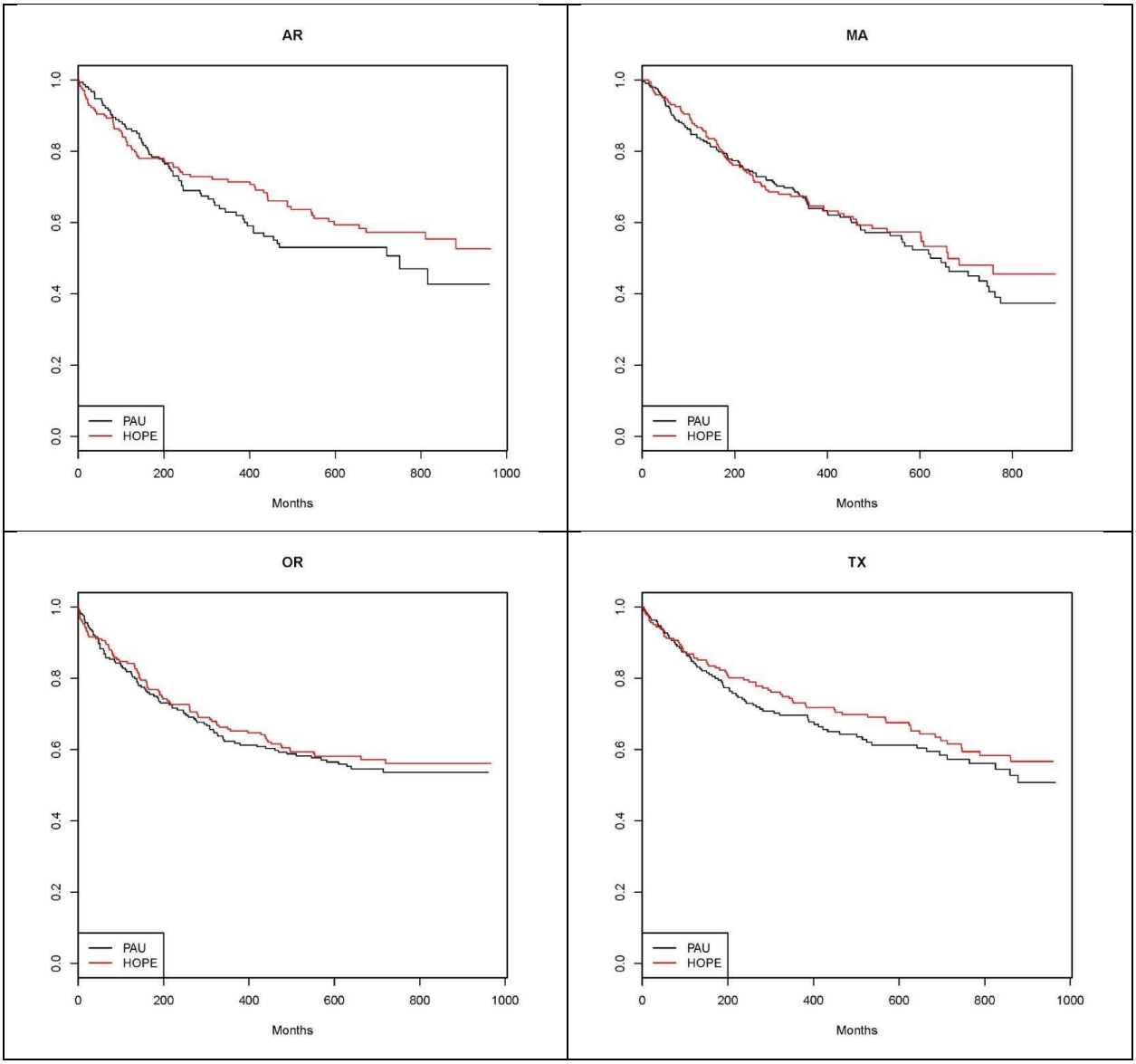


Exhibit 5-8 shows log-normal survival model results for time to first arrest.[32] For log-normal survival models, the coefficient estimates are the impact of the variable on the mean survival time—so positive coefficients indicate that individuals with that trait (or more of that trait) have longer times to the event of interest. The first models include only the HOPE indicator and for the overall model site indicators. These results show that ***HOPE program participation was not significantly related to time to first rearrest for any of the four sites. Time to first arrest was less for Oregon participants than for the reference site (Texas) participants (p = 0.05) in the overall model.***

[33] Observations were censored on death and end of the follow-up period.

Exhibit 5-8. Log-normal survival model results for time to first new arrest.

Parameter	Overall Estimate[1] (SE)	AR Estimate (SE)	MA Estimate (SE)	OR Estimate (SE)	TX Estimate (SE)
Arrest Model 1					
Intercept	**7.02*** (0.15)**	**6.74*** (0.21)**	**6.45*** (0.14)**	**10.25*** (0.59)**	**7.01*** (0.21)**
Group: HOPE = 1	0.14 (0.13)	0.12 (0.27)	0.11 (0.19)	0.12 (0.30)	0.21 (0.27)
Site = AR	-0.25 (0.19)				
Site = MA	-0.29 (0.18)				
Site = OR	**-0.35* (0.18)**				
Chisq	5.57	0.18	0.32	0.16	0.63
Arrest Model 2					
Intercept	**7.45*** (0.24)**	**6.87*** (0.37)**	**6.66*** (0.38)**	**7.42*** (0.49)**	**8.06*** (0.44)**
Group: HOPE = 1	-0.21 (0.29)	0.37 (0.54)	-0.31 (0.54)	-0.33 (0.68)	-0.72 (0.57)
Sex: Male = 1	**-0.61** (0.23)**	-0.22 (0.42)	-0.25 (0.40)	-0.56 (0.52)	**-1.37** (0.47)**
Group x Sex	0.41 (0.32)	-0.30 (0.62)	0.47 (0.57)	0.48 (0.65)	1.15 (0.64)
Site = AR	-0.26 (0.18)				
Site = MA	-0.24 (0.17)				
Site = OR	-0.29 (0.17)				
Chisq	13.49	1.82	0.98	1.79	**9.76***
N	1477	321	391	394	371

[1] Texas is the reference category for the site indicators in the overall models.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$

The second set of models includes, in addition to the HOPE indicator, a sex indicator and an interaction between group and sex to assess whether HOPE has differential effects on males and females. ***There was no main effect of HOPE when controlling for sex and the sex by HOPE intervention. Males had significantly shorter times to first arrest in the overall model (p = 0.007) and in the Texas model (p = 0.004).*** Males in the HOPE group in Texas had marginally longer times to first arrest than control group males when controlling for the gender and intervention main effects (p = 0.07). Thus, with respect to the effect of HOPE on arrest, the survival results are consistent with the bivariate results discussed earlier.

Exhibit 5-9 shows the Kaplan-Meier survival curves for the HOPE and PAU groups (overall) for time to first revocation.[33] Here, in contrast to the results for new arrests, the red HOPE line sits largely below the black PAU line, suggesting shorter times to revocation for HOPE probationers compared with PAU; however, again, the difference is not significant.

[33] Observations were censored on death and end of the follow-up period.

Exhibit 5-9. Kaplan Meier survival curves for time to first revocation for all HOPE and PAU probationers

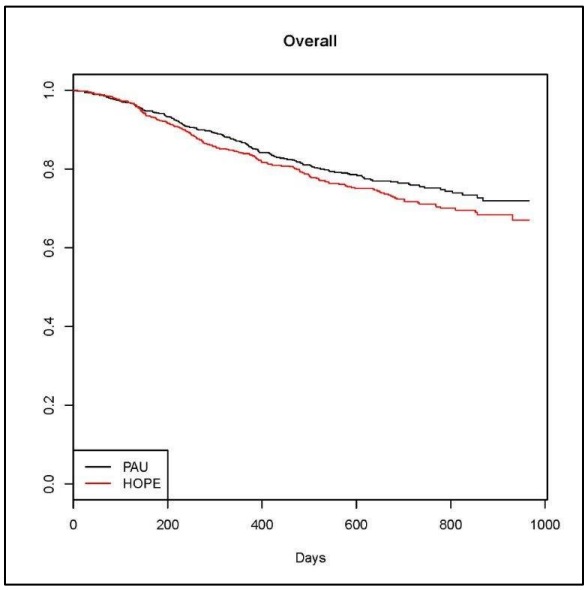
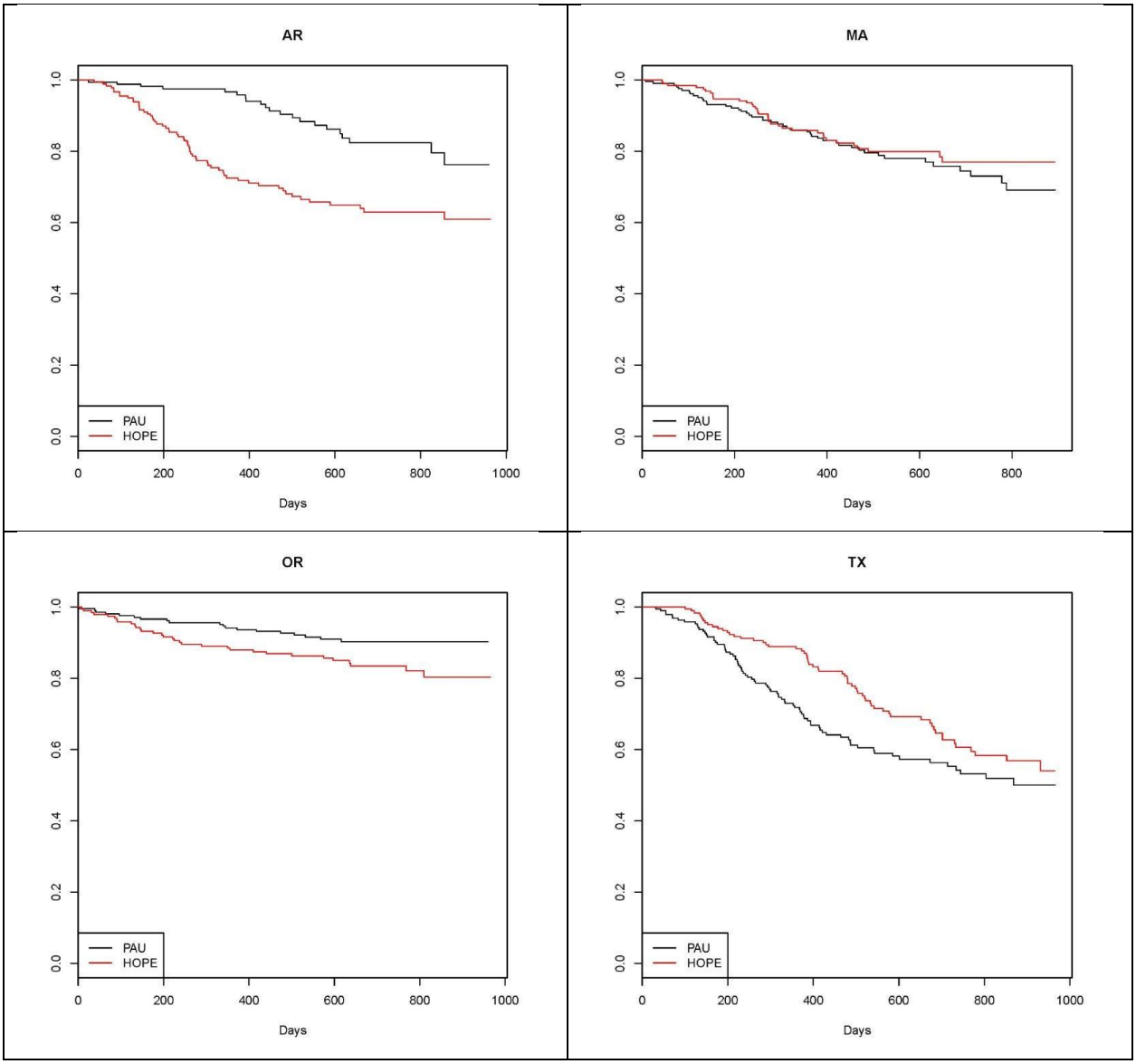


Exhibit 5-10 shows the survival curves for time to first revocation for the HOPE and PAU groups by site. As can be seen, there are across-site differences, with the time to first revocation similar for the two groups in Massachusetts but divergent in the other three sites***. HOPE participants were revoked more quickly in Arkansas and Oregon and more slowly in Texas.***

Exhibit 5-11 shows log-normal survival model results for time to revocation. As before, Model 1 includes only the HOPE indicator and, for the overall model, site indicators***. The coefficient estimate for the HOPE indicator was not significant in the overall model*** (p = 0.14), likely because of the differential effects of HOPE participation on revocation across the sites (as visible in ***Exhibit 5-10***). These differential effects are evident in the site-level models***. HOPE program participants experienced shorter times on average to revocation than did those on PAU in Arkansas and Oregon and longer times on average in Texas; there was no program effect in Massachusetts.*** The second model includes the sex indicator; however, we excluded the interaction term between HOPE and sex in these models because of substantial collinearity between the interaction term and the sex variable. In the overall model, males are revoked more quickly than females, with the remaining results the same as in Model 1. In the site-level models, the results for HOPE participation are the same as in Model 1—***HOPE participants were revoked more quickly in Arkansas and Oregon and less quickly in Texas.*** Males had shorter times to revocations in all sites, but these results were statistically significant only in and marginally so in Arkansas (p = 0.06).

Exhibit 5-10. Kaplan-Meier survival curves for time to first new arrest (charge) by site and group



Results for Model 2 show ***shorter times to revocations for male subjects***—significantly so for the overall model and the Texas model. The group x sex interaction was dropped from this model because of collinearity.

Exhibit 5-11. Log-normal survival model results for time to first revocation

Parameter	Overall Estimate[1] (SE)	AR Estimate (SE)	MA Estimate (SE)	OR Estimate (SE)	TX Estimate (SE)
Revoked Model 1					
Intercept	**7.66*** (0.15)**	**8.00*** (0.24)**	**7.58*** (0.21)**	**10.25*** (0.59)**	**6.63*** (0.11)**
Group: HOPE = 1	**–**0.17 (0.11)	**–1.02*** (0.23)**	0.17 (0.23)	**–0.94* (0.46)**	**0.41** (0.14)**
Site = MA	0.06 (0.16)				
Site = OR	**0.65*** (0.17)**				
Site = TX	**-0.44** (0.16)**				
Chisq	50.65***	22.19***	0.56	4.49*	8.43**
Revoked Model 2					
Intercept	**8.04*** (0.20)**	**8.34***(0.31)**	**7.81*** (0.51)**	**10.51*** (0.79)**	**7.04*** (0.18)**
Group: HOPE = 1	-0.17 (0.11)	**-0.98*** (0.23)**	0.17 (0.23)	**-0.94* (0.46)**	**0.40** (0.14)**
Sex: Male = 1	**-0.50***(0.15)**	-0.49 (0.26)	-0.32 (0.38)	-0.32 (0.61)	**-0.52** (0.18)**
Group x Sex[2]					
Site = MA	0.12 (0.16)				
Site = OR	**0.68*** (0.17)**				
Site = TX	**-0.43** (0.16)**				
Chisq	62.15***	25.91***	1.33	4.76	**17.44*** **
N	1496	340	391	394	371

[1] Arkansas is the reference category for the site indicators in the overall models.
[2]Interaction term was excluded from the models because of collinearity.
*p < 0.05. **p < 0.01. ***p < 0.001.

Results for time to first arrest or revocation (whichever occurred first) were similar to those for the arrest models—no significant differences between those on HOPE and PAU. Exhibit 5-12 shows survival curves for time to revocation or arrest[34] for the HOPE and PAU groups combined. As can be seen, there is little difference between the two curves, suggesting that overall there were no differences in time to first arrest or revocation.

[34] Observations were censored on death and end of the follow-up period.

Exhibit 5-12. Kaplan Meier survival curves for time to first revocation for all HOPE and PAU probationers.

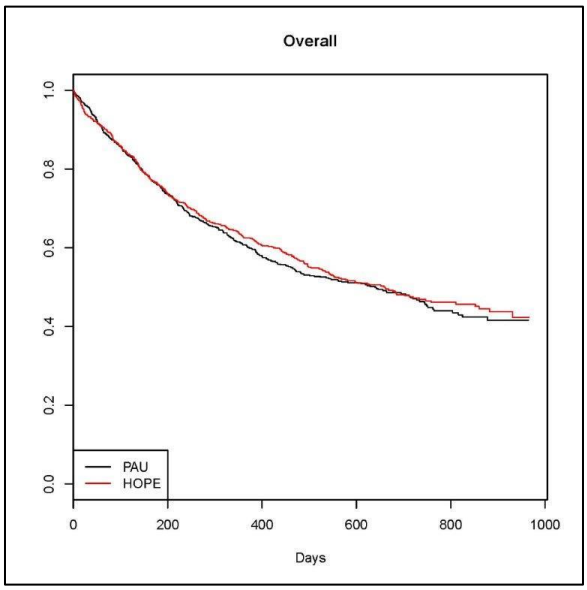


Exhibit 5-13 shows the survival curves for time to first arrest or revocation for the HOPE and PAU groups by site. As can be seen, ***the greatest difference between the HOPE and PAU curves is apparent in the Texas graph—with HOPE probationers having somewhat longer times to experiencing either a first arrest or revocation.***

Exhibit 5-13. Kaplan-Meier survival curves for time to first new arrest (charge) or revocation by site and group

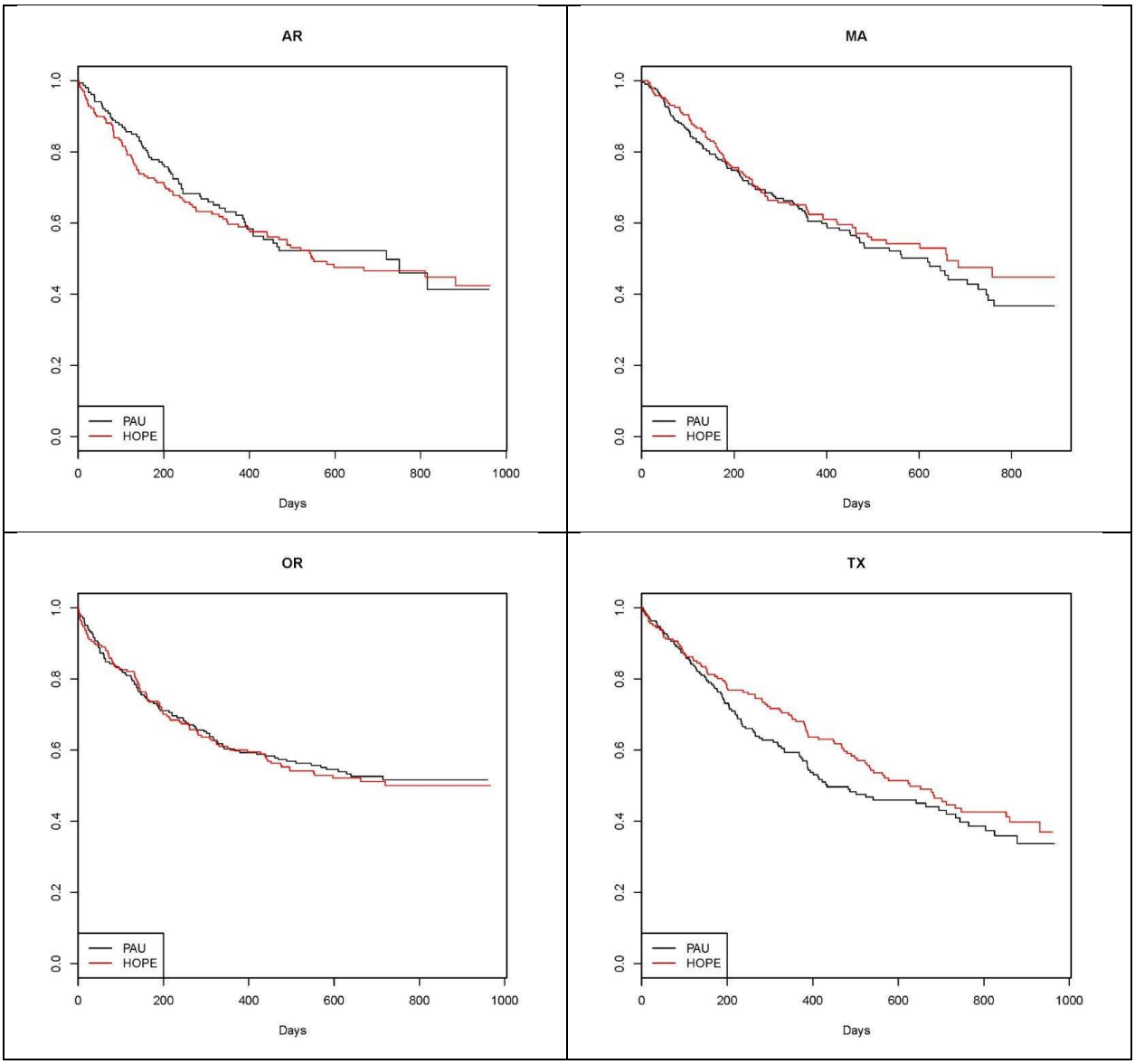


Exhibit 5-14 shows the log-normal survival results for time to arrest or revocation. As expected from the graphs in Exhibit 5-13, ***there is no difference between HOPE or PAU groups overall or within any of the sites.*** Model 2 once again shows shorter times to rearrest for male participants—significantly shorter overall and in Texas.

Exhibit 5-14. Log-normal survival model results for time to first arrest or revocation

Parameter	Overall Estimate[1] (SE)	AR Estimate (SE)	MA Estimate (SE)	OR Estimate (SE)	TX Estimate (SE)
Arrest/Rev Model 1					
Intercept	**6.44***(0.12)**	**6.58***(0.19)**	**6.33*** (0.13)**	**6.71*** (0.20)**	**6.26*** (0.14)**
Group: HOPE = 1	0.03 (0.11)	-0.24 (90.24)	0.14 (0.18)	-0.08 (0.26)	0.21 (0.19)
Site = AR	-0.01 (0.16)				
Site = MA	0.10 (0.15)				
Site = OR	-0.02 (0.15)				
Chisq	0.84	1.02	0.60	0.10	1.19
Arrest/Rev Model 2					
Intercept	**6.77***(0.17)**	**6.91*** (0.28)**	**6.25*** (0.27)**	**7.24*** (0.37)**	**6.74*** (0.23)**
Group: HOPE = 1	0.03 (0.11)	-0.21 (0.24)	0.14 (0.18)	-0.08 (0.26)	0.19 (0.19)
Sex: Male = 1	**-0.46** (0.14)**	-0.46 (0.28)	0.08 (0.27)	-0.64 (0.36)	**-0.63** (0.23)**
Group x Sex[2]					
Site = MA	0.16 (0.16).				
Site = OR	0.02 (0.16)				
Site = TX	0.00 (0.16)				
Chisq	11.24*	3.82	0.68	3.25	**9.03***
N	1477	321	391	394	371

[1] Arkansas is the reference category for the site indicators in the overall models.
[2] The group x sex term was dropped from the model because of collinearity.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

The next set of recidivism analyses examines the time to a new conviction. ***Exhibit 5-15*** shows survival curves for the HOPE and PAU groups combined. As can be seen, there is little difference between the two curves, suggesting that ***overall there were no differences between groups in time to a new conviction.***

Exhibit 5-15. Kaplan Meier survival curves for time to first conviction for all HOPE and PAU probationers

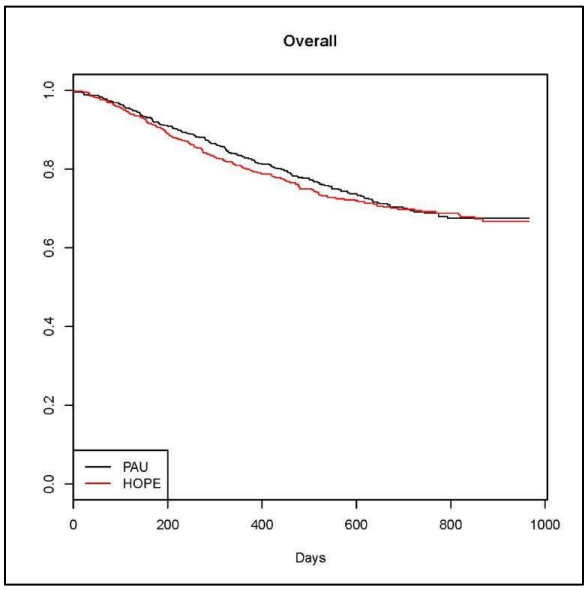


Exhibit 5-16 shows the survival curves for time to first new conviction for the HOPE and PAU groups by site. As can be seen, ***the greatest difference between the HOPE and PAU curves is apparent in the Arkansas graph—with HOPE probationers experiencing revocation much sooner than the PAU probationers.*** The was no difference in Massachusetts and no significant differences in Oregon and Texas in time to first new conviction.

The log-normal survival results are shown in ***Exhibit 5-17***. In three of the four sites, there was no statistically significant difference between the HOPE and PAU groups. ***In Arkansas, HOPE program participants experienced shorter times to a new conviction.*** Model 2 results once again suggest shorter times to new convictions for males.

So, the conclusions of the survival models confirm the results from the bivariate comparisons reported in the previous section—no differences except for worse outcomes for HOPE participants on revocations in Arkansas and Oregon and on reconvictions in Arkansas and better outcomes on time to revocation for HOPE in Texas.

Exhibit 5-16. Kaplan-Meier survival curves for time to first new conviction by site and group

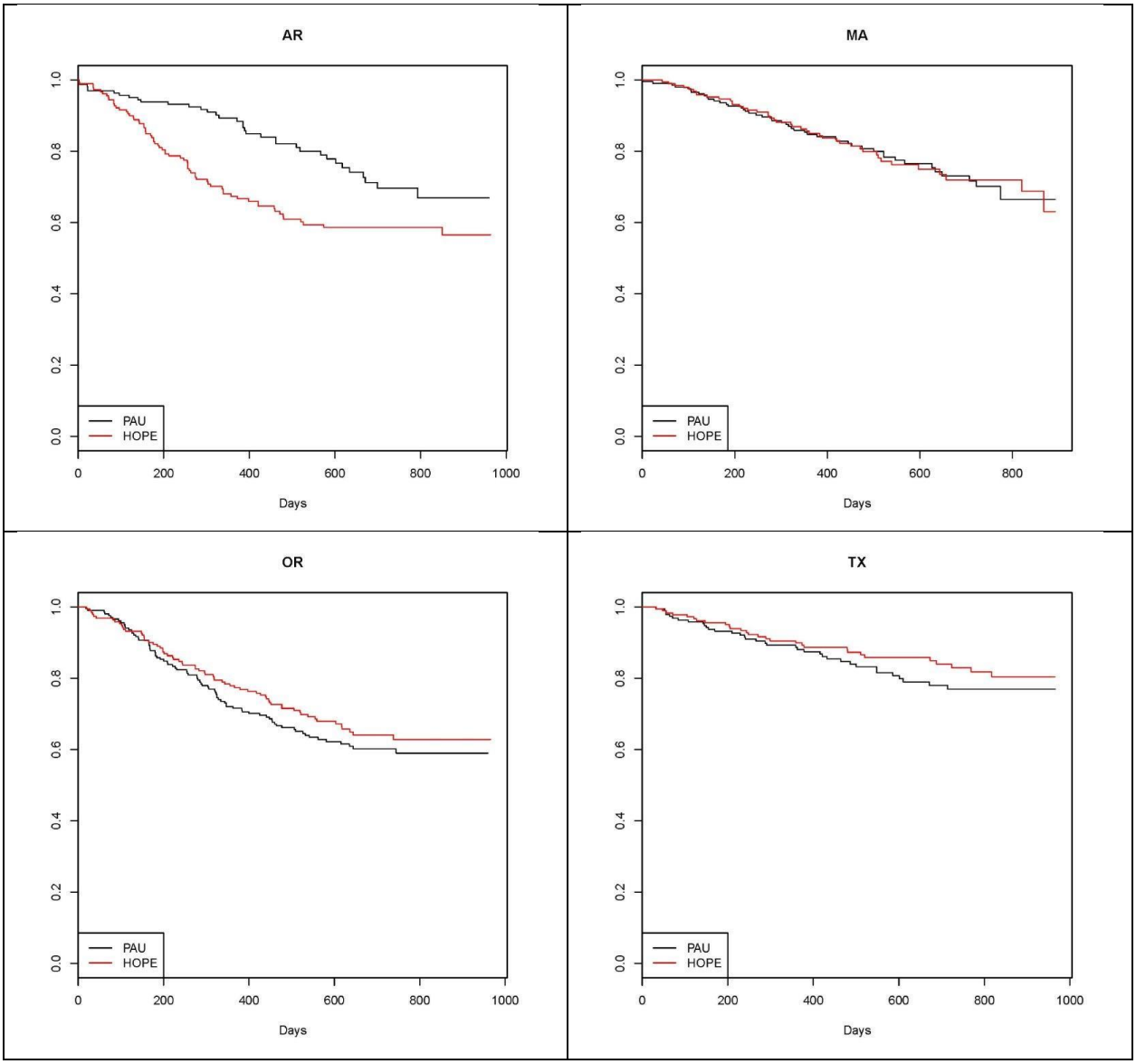


Exhibit 5-17. Log-normal survival model results for time to first new conviction

Parameter	Overall Estimate[1] (SE)	AR Estimate (SE)	MA Estimate (SE)	OR Estimate (SE)	TX Estimate (SE)
Conviction Model 1					
Intercept	**7.24*** (0.15)**	**7.92*** (0.27)**	**7.49*** (0.19)**	**6.97*** (0.14)**	**8.14*** (0.29)**
Group: HOPE = 1	-0.07 (0.12)	**–0.88*** (0.28)**	0.06 (0.21)	0.15 (0.18)	0.27 (0.29)
Site = MA	**0.48** (0.17)**				
Site = OR	-0.03 (0.16)				
Site = TX	**0.88*** (0.18)**				
Chisq	**39.77*** **	**9.99*** **	0.08	0.63	0.93
Conviction Model 2					
Intercept	**7.60*** (0.19)**	**8.23*** (0.36)**	**7.61*** (0.36)**	**7.27***(0.26)**	**8.85*** (0.45)**
Group: HOPE = 1	-0.07 (0.12)	**-0.84** (0.28)**	0.06 (0.21)	0.14 (0.18)	0.22 (0.28)
Sex: Male = 1	**-0.49** (0.16)**	-0.45 (0.32)	-0.12 (0.34)	-0.36 (0.25)	**-0.91 (0.38)***
Group x Sex					
Site = MA	**0.54** (0.17)**				
Site = OR	0.02 (0.16)				
Site = TX	**0.89*** (0.18)**				
Chisq	**49.83*** **	**11.98** **	0.21	2.75	**7.41***
N	1496	340	391	394	371

[1] Arkansas is the reference category for the site indicators in the overall models.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Recidivism: Competing Hazards Model Results

Exhibit 5-18 shows the Kaplan–Meyer cumulative failure curves for time to first arrest by offense type (person, property, drug, and public order/other) and study assignment for the total sample. As can be seen, most dashed and solid lines of the same color track closely—***suggesting no differences between the HOPE and PAU groups. The most substantial difference is for failure by a drug-related arrest, with the PAU group failing more quickly on a drug charge when accounting for failure by other offense types.***

Exhibit 5-19 shows the results by site and group. There are few observable differences in these curves. The exception is ***the difference between failure for a drug-related arrest in Texas, with substantially longer time to failure for a drug-related offense for the HOPE group,*** which appears to be driving the difference observed across all samples.

Exhibit 5-18. Kaplan Meier survival curves for time to first arrest by offense type for all HOPE and PAU probationers

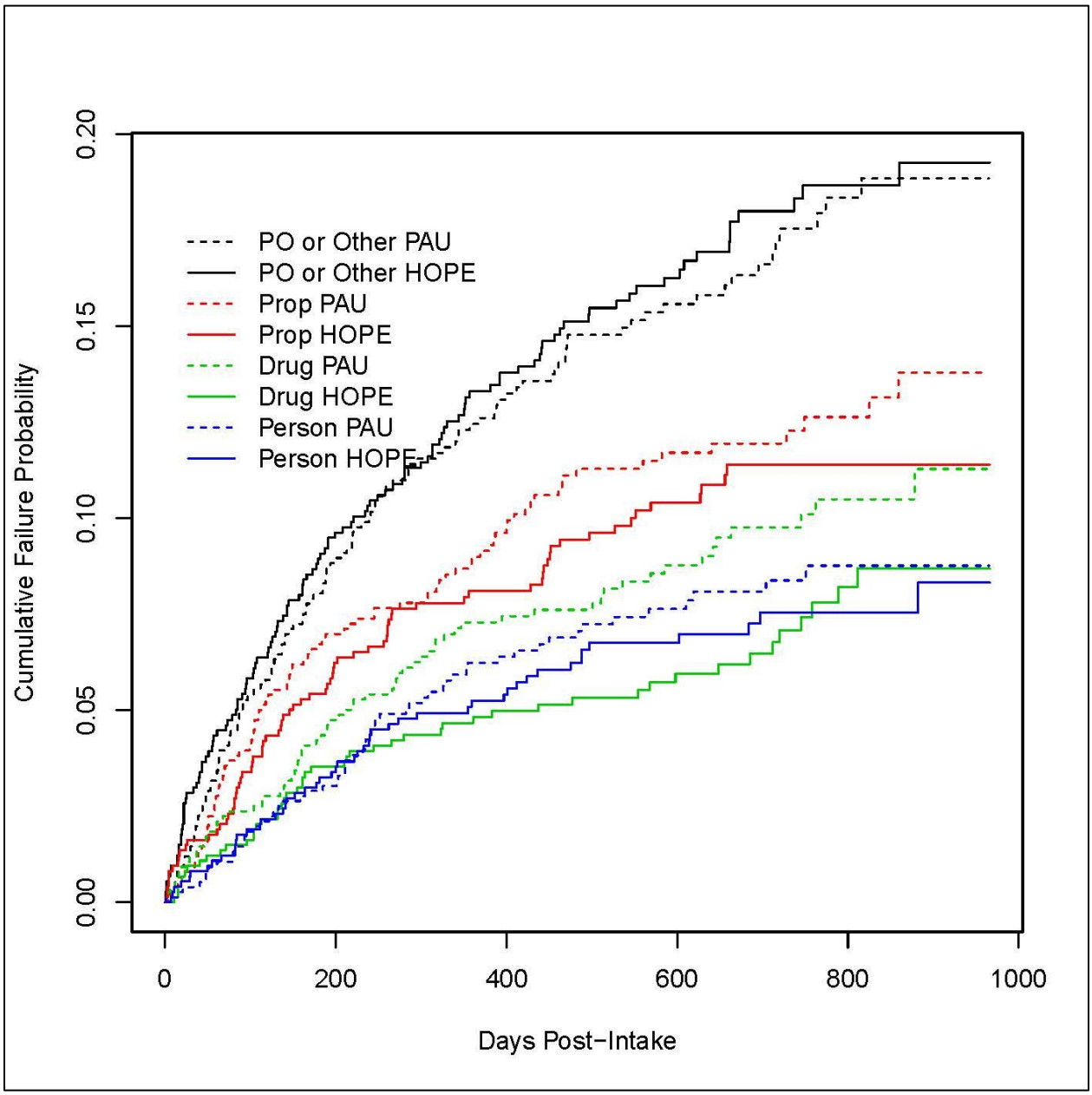
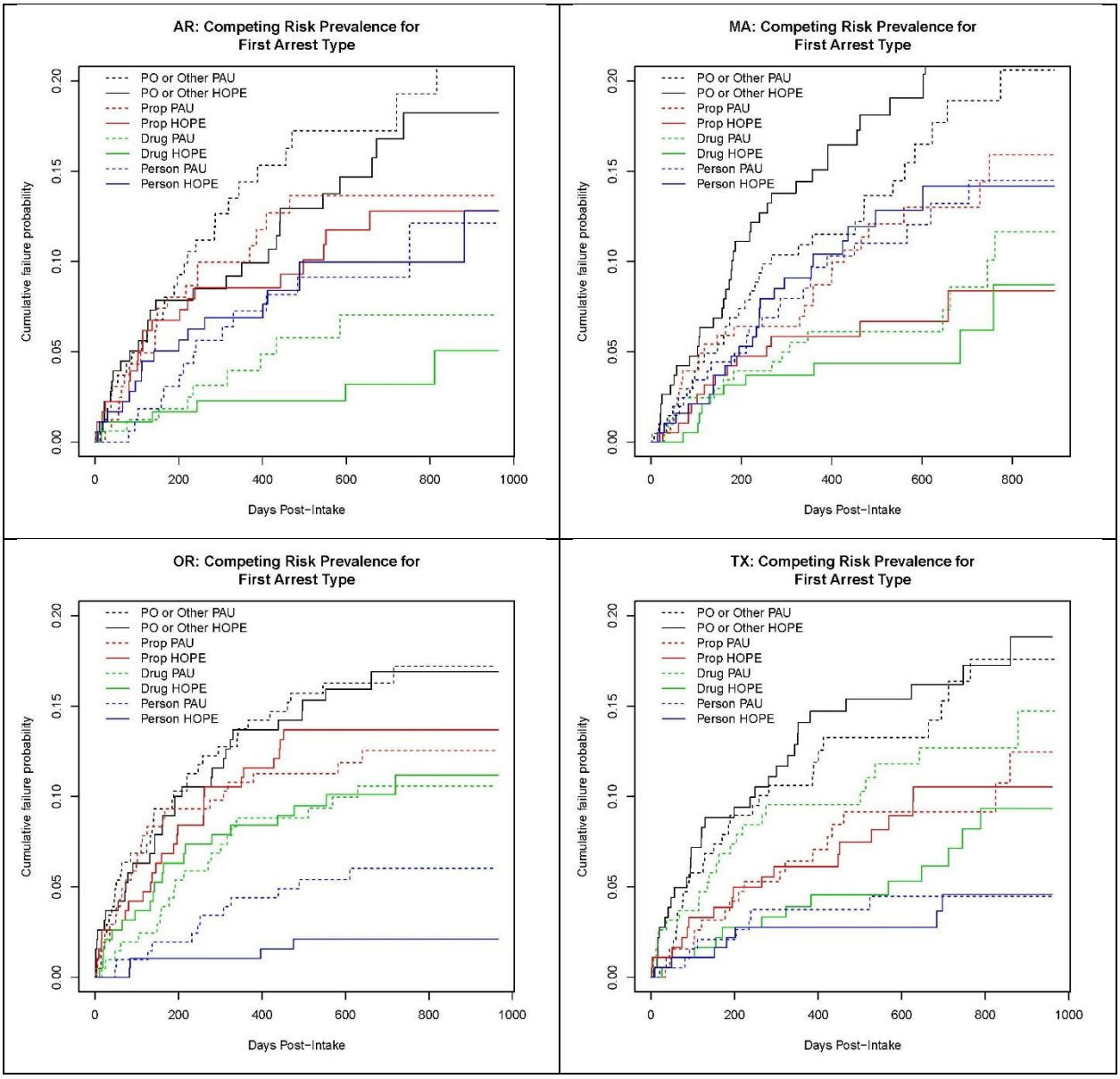


Exhibit 5-19. Kaplan-Meier survival curves for time to first arrest by offense type, site, and group



Results for the competing hazard Cox regression models are shown in ***Exhibit 5-20***. (A negative coefficient implies a lower hazard rate—or less risk—for individuals in the HOPE group.) The only coefficient value for the overall model that approaches significance is the value for drug offense with an estimate for the HOPE coefficient of –0.35 and a standard error of 0.19 (exp(–0.35) = 0.706), suggesting lower risk for HOPE participants of a drug-related risk at $p = 0.06$ (z score = –1.862). ***Exhibit 5-20*** also shows the results for each competing event by site. ***There are no differences at the p < 0.05 level between the HOPE and PAU groups.*** The result for Oregon for person offenses is significant at the 0.06 level (z = -1.817), with HOPE participants faring somewhat better than the PAU participants (i.e., lower risk of arrest for a person offense). The result for Texas for drug offenses is significant at the 0.08 level (z = -1.741), with again HOPE participants faring somewhat better than the PAU participants.

Exhibit 5-20 Cox regression competing risk model results: first arrest for person charge, property charge, drug charge, public order, or other charge

First Offense	Overall Estimate (SE)	AR Estimate (SE)	MA Estimate (SE)	OR Estimate (SE)	TX Estimate (SE)
Person Group: HOPE = 1	-0.14 (0.19)	0.01 (0.36)	0.05 (0.29)	–1.05 (0.58)	–0.16 (0.52)
Property Group: HOPE = 1	-0.14 (0.16)	–0.09 (0.31)	–0.55 (0.34)	0.09 (0.28)	–0.10 (0.34)
Drug Group: HOPE = 1	-0.35 (0.19)	–0.63 (0.53)	–0.34 (0.40)	0.01 (0.31)	–0.60 (0.35)
Public Order or Other Group: HOPE = 1	-0.02 (0.13)	–0.33 (0.28)	0.19 (0.24)	–0.04 (0.25)	0.05 (0.26)
N	1496	340	391	394	371

Results are not significant at the $p < 0.05$ level.

Recidivism: Negative Binomial Count Models

In addition to any event, time to event, and time to competing events, we also looked at the number of recidivism events using negative binomial count models. (We only had reliable arrest charge data for Arkansas, Oregon, and Texas; data for Massachusetts carried the potential for overcounting events so were excluded.)[35] ***Exhibit 5-21*** shows the results for the overall model and the site-specific models for number of arrests. The number of arrests ranged from 0 to 15. ***The parameter estimates for the HOPE indicator are not significantly different from zero (at p < 0.05) for any of the models.*** Males have higher arrest counts than females overall and in Texas (which, again, likely is driving the overall results).

Exhibit 5-21. Negative binomial model results for number of new arrests

Parameter	Overall Estimate[1] (SE)	AR Estimate (SE)	OR Estimate (SE)	TX Estimate (SE)
Intercept	**-6.64*** (0.21)**	**-6.00*** (0.30)**	**-6.58*** (0.33)**	**-7.16*** (0.36)**
Group: HOPE = 1	-0.19 (0.27)	-0.55 (0.47)	0.06 (0.48)	0.17 (0.49)
Sex: Male = 1	**0.50* (0.21)**	0.05 (0.36)	0.39 (0.36)	****1.28 (0.40)**
Group x Sex	0.04 (0.31)	0.65 (0.55)	-0.03 (0.52)	-0.78 (0.55)
Site = AR	**0.34* (0.16)**			
Site = OR	0.07 (0.15)			
N	1103	338	394	371

Texas is the reference category for the site indicators in the overall model; data were unavailable for Massachusetts.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

[35]The Massachusetts data were arraignment data; cases that were appealed or that were "continued without finding" could appear as new arraignment records in the data although how they were handled was inconsistent.

Exhibit 5-22 shows the results for the overall model and the site-specific models for number of arrest charges. Values ranged from 0 to 30. ***The parameter estimates for the HOPE indicator are not significantly different from zero (at p < 0.05) overall and for the Oregon and Texas models; HOPE probationers had fewer arrest charges than PAU in Arkansas.*** Males have a higher number of arrest charges than females overall and in Texas (which, again, likely is driving the overall results).

Exhibit 5-22. Negative binomial model results for number of new arrest charges

Parameter	Overall Estimate[1] (SE)	AR Estimate (SE)	OR Estimate (SE)	TX Estimate (SE)
Intercept	**-6.14*** (0.26)**	**-4.83*** (0.36)**	**-5.90*** (0.42)**	**-7.02*** (0.39)**
Group: HOPE = 1	-0.59 (0.33)	**-1.27* (0.56)**	-0.07 (0.60)	0.14 (0.54)
Sex: Male = 1	**0.49* (0.26)**	-0.35 (0.43)	0.62 (0.45)	**1.68*** (0.44)**
Group x Sex	0.48 (0.37)	**1.49* (0.65)**	0.28 (0.66)	-0.84 (0.61)
Site = AR	**0.76*** (0.19)**			
Site = OR	**0.52** (0.18)**			
N	1103	338	394	371

Texas is the reference category for the site indicators in the overall model; data were unavailable for Massachusetts.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Exhibit 5-23 shows the results for the overall model and the site-specific models for number of recidivism convictions. Values ranged from 0 to 5. ***The parameter estimates for the HOPE are not significantly different from zero (at p < 0.05) overall and for the Arkansas and Oregon models; males have higher numbers of new convictions than females. In Texas, HOPE probationers had marginally fewer new convictions than PAU (p = 0.09), males had fewer convictions than females (p = 0.009) and HOPE males had fewer convictions than PAU males (p = 0.035).***

Exhibit 5-23. Negative binomial model results for number of new convictions

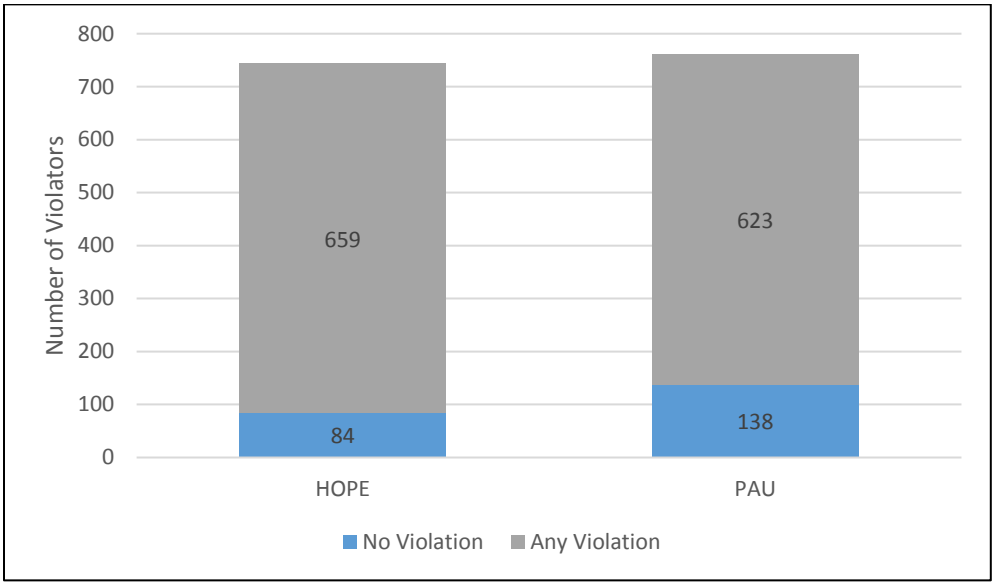
Parameter	Overall Estimate[1] (SE)	AR Estimate (SE)	OR Estimate (SE)	TX Estimate (SE)
Intercept	**-8.06*** (0.27)**	**-6.89*** (0.38)**	**-7.10*** (0.36)**	**-9.20*** (0.76)**
Group: HOPE = 1	0.15 (0.33)	-0.21 (0.55)	-0.03 (0.53)	1.52 (0.88)
Sex: Male = 1	**0.51* (0.26)**	-0.40 (0.45)	0.45 (0.38)	**2.10** (0.80)**
Group x Sex	0.15 (0.37)	**1.57* (0.64)**	0.12 (0.57)	**-2.00* (0.95)**
Site = AR	**0.88*** (0.20)**			
Site = OR	**0.85*** (0.18)**			
N	1103	338	394	371

[1] Texas is the reference category for the site indicators in the overall model; data were unavailable for Massachusetts.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

5.2. Impact of HOPE Participation on Supervision Compliance

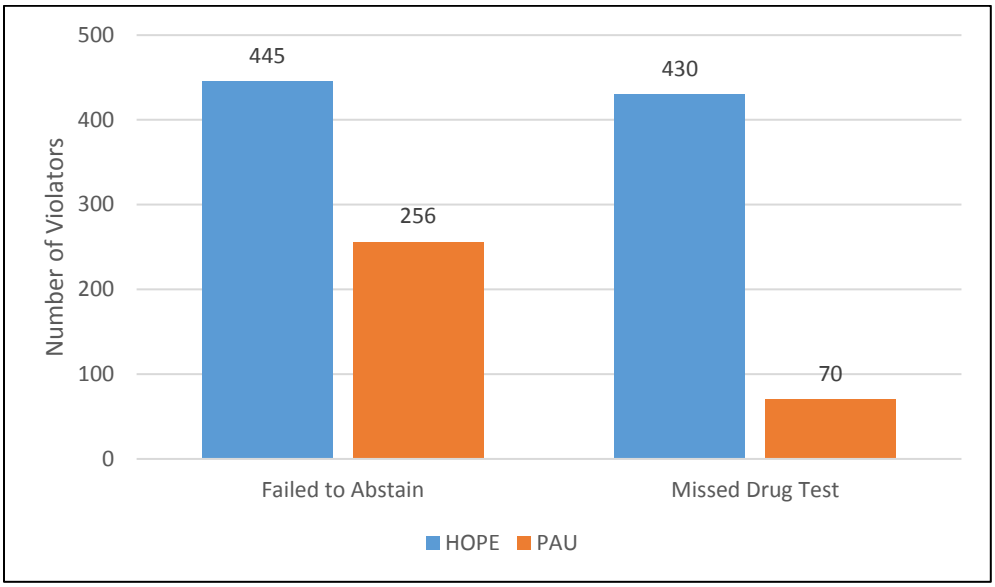
Swift and certain responses for any violations are hypothesized by the HOPE model to lead to better supervision compliance. Examining this issue is complicated by the fact that the increased scrutiny associated with HOPE is almost certain to increase the likelihood that violations will be detected and, because of the requirements of HOPE, reported and responded to. ***Exhibit 5-24*** shows ***that HOPE probationers were more likely than PAU probationers to have had at least one probation violation*** (chi-square = 13.93, p = 0.0002), although most individuals in both groups had at least one violation.

Exhibit 5-24. Numbers of probation violators by group



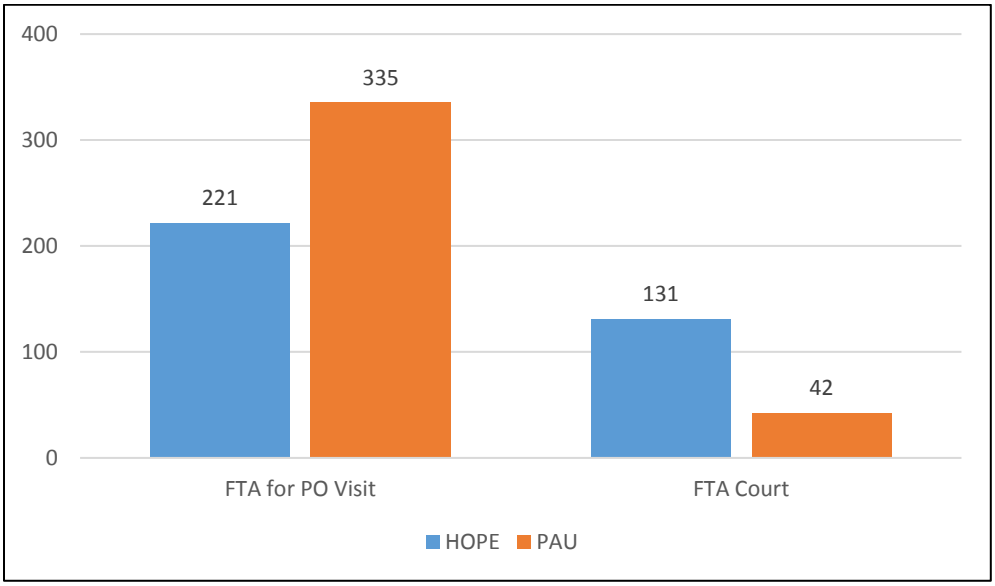
Many violations were related to mandatory random drug testing for HOPE. ***Exhibit 5-25*** shows the number of each group who had at least one violation for failing to abstain from drug use and missing at least one drug test. ***HOPE probationers were significantly more likely to have had a violation for drug use*** (chi-square = 104.12, p < 0.0001) ***and for missing a drug test*** (chi-square = 401.36, p < 0.0001). Overall, 60% of the HOPE group and 34% of the PAU group had at least one violation for failing to abstain from drug use, while 58% of HOPE probationers compared to 9% of PAU probationers had missed at least one drug test.

Exhibit 5-25. Numbers of violators for drug-related issues by group



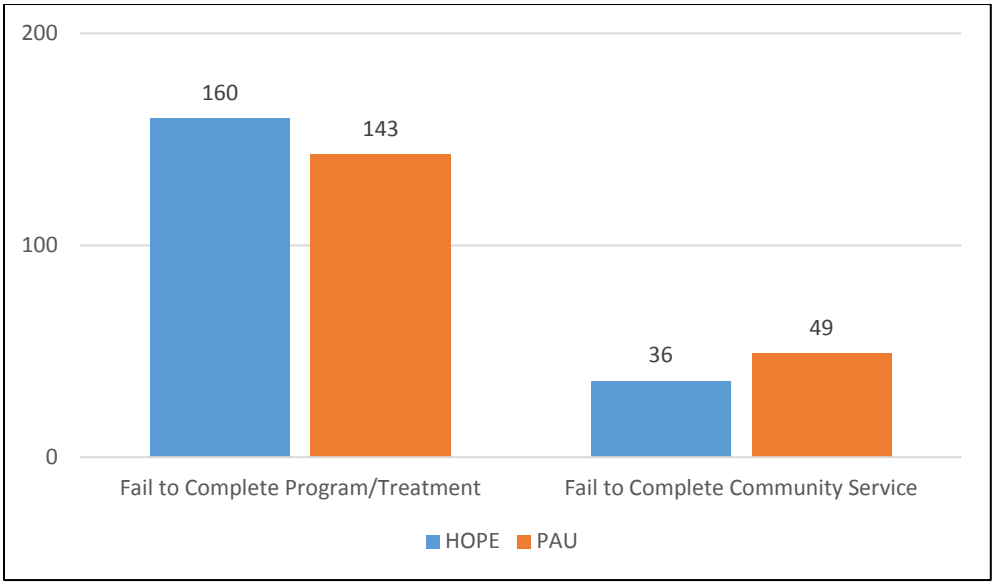
Failures to appear were less frequent for HOPE probationers than drug-related violations (***Exhibit 5-26)***. Overall, 30% of the HOPE group compared to 44% of the PAU group had at least one violation for failing to appear for a probationer office visit; 18% of HOPE compared to 6% of HOPE had a violation for failing to appear in court. ***Those on PAU were significantly more likely than those on HOPE probation to receive a violation for missing an appointment with a probation officer*** (chi-square = 32.88, $p < 0.0001$), while ***those on HOPE probation were more likely to fail to appear for court*** (chi-square = 54.18, $p < 0.0001$). These results may simply reflect the fact that the HOPE group had more appearances and, thus, increased opportunity to miss an appearance.

Exhibit 5-26. Numbers of violators for failures to appear by group



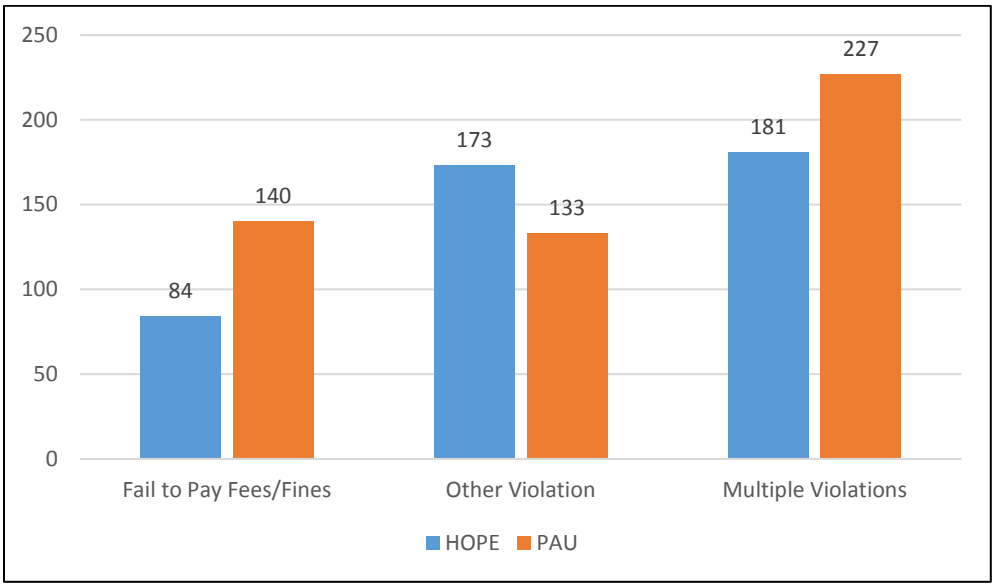
Some individuals also received violations for failing to complete programs or treatment and for failing to complete community service (***Exhibit 5-27)***. The numbers who experienced violations for these conditions were much smaller than for the previously discussed violations. ***There were no differences between the HOPE and PAU groups on these two measures.*** About 20% of both groups received violations for failing to complete programs or treatment (chi-square = 1.76, p = 0.18), while roughly 5% of both groups failed to complete community service (chi-square = 1.79, p < 0.18).

Exhibit 5-27. Numbers of violators for failing to complete programs/treatment or community service



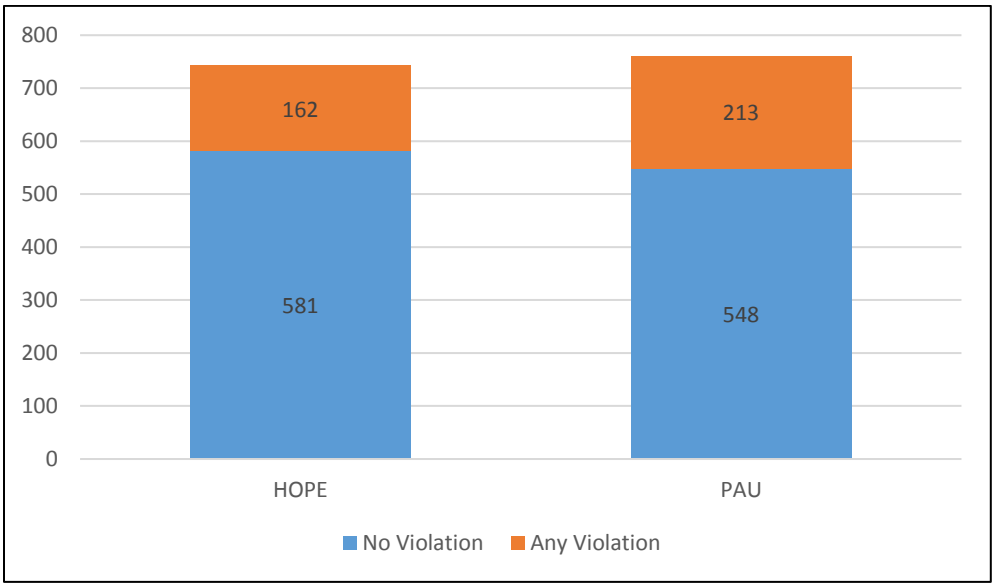
Some individuals received violations for failing to pay their fines or fees, for some violation other than those previously discussed, or for multiple violations for the same incident (***Exhibit 5-28)***. ***Those on PAU (18%) were much more likely to have received a violation for failing to pay fees or fines than those on HOPE*** (11%; chi-square = 14.92, p = 0.0001). ***Those on HOPE (23%) were much more likely to have received a violation for something other than those previously identified than those on PAU*** (17%; chi-square = 7.82, p = 0.005). ***Those on PAU (30%) were more likely to receive a violation with multiple charges than those on HOPE*** (24%; chi-square = 5.69, p = 0.017).

Exhibit 5-28. Numbers of violators for failing to pay fines or fees or for another violation



There was also a significant difference in the numbers who received violations for new arrests/charges. As can be seen in ***Exhibit 5-29***, ***those on PAU*** (28%) ***were much more likely to have received a violation for a new charge than those on HOPE*** (22%; chi-square = 7.68, p = 0.006).[36]

Exhibit 5-29. Numbers of probation violators for a new arrest/charge by group



[36] This measure reflects having a probation violation for a new arrest or charge. Results presented in Exhibit 5-1 showed no differences between HOPE and PAU in actually having any arrest charge (40% of HOPE and 44% of PAU; ns), although HOPE probationers were less likely than PAU to have a property or drug charge. Individuals who are arrested may not incur a violation specifically for an arrest, so these measure two different constructs. Exhibit 5-2 compared arrest measures—any or for specific charges—for HOPE and PAU groups by site—the only significant finding was a lower likelihood of a drug charge for HOPE probationers in Texas.

Exhibit 5-30 shows the mean numbers of violations overall and by type by group. Also shown are the maximum values for the two groups for each type. (The minimum counts for all types and both groups were zero.) As can be seen, most differences are significant and in most cases the difference favors those on PAU. On average, ***HOPE probationers had slightly more than five violations compared to slightly more than four for those on PAU***. This means that the 743 HOPE probationers had a total of 3,770 violations compared to 3,134 for the 761 PAU probationers. The number of violations ranged from zero to 25 for one or more HOPE probationers and to 29 for one or more PAU probationers. ***HOPE probationers had more than twice as many violations for failing to abstain from drug use and nearly ten times the number of violations for missing drug tests.*** These findings are not surprising as HOPE probationers were exposed to much more substantial drug testing requirements, making it more likely that their drug use would be detected and more likely that they would miss a test.

Exhibit 5-30. Means (and standard deviations) of counts of probation violations by group

Type of Violation	HOPE	PAU	t Value	HOPE Maximum	PAU Maximum
Any violation‡	**5.07‡ (4.43)**	**4.12 (4.40)**	-4.2	25	29
Fail to abstain from drug use‡	**1.42‡ (1.76)**	**0.65 (1.32)**	-9.65	10	14
Missed drug test‡	**1.42‡ (1.76)**	**0.15 (0.58)**	-18.63	9	6
Fail to appear PO visit‡	**0.43‡ (0.82)**	**1.14 (1.92)**	9.41	9	18
Fail to appear court‡	**0.23‡ (0.59)**	**0.06 (0.26)**	-7.17	6	2
Fail to complete program/treatment	0.30 (0.68)	0.29 (0.73)	-0.34	6	6
Fail to complete community service†	**0.06† (0.28)**	**0.10 (0.46)**	2.34	3	5
Fail to pay fees/fines‡	**0.18‡ (0.59)**	**0.39 (1.21)**	4.27	6	11
Other type of violation‡	**0.33‡ (0.74)**	**0.22 (0.57)**	-3.16	7	6
Multiple causes for single violation‡	**0.40‡ (0.86)**	**0.63 (1.34)**	3.97	5	9
New charge‡	**0.31‡ (0.68)**	**0.48 (1.07)**	3.78	5	9

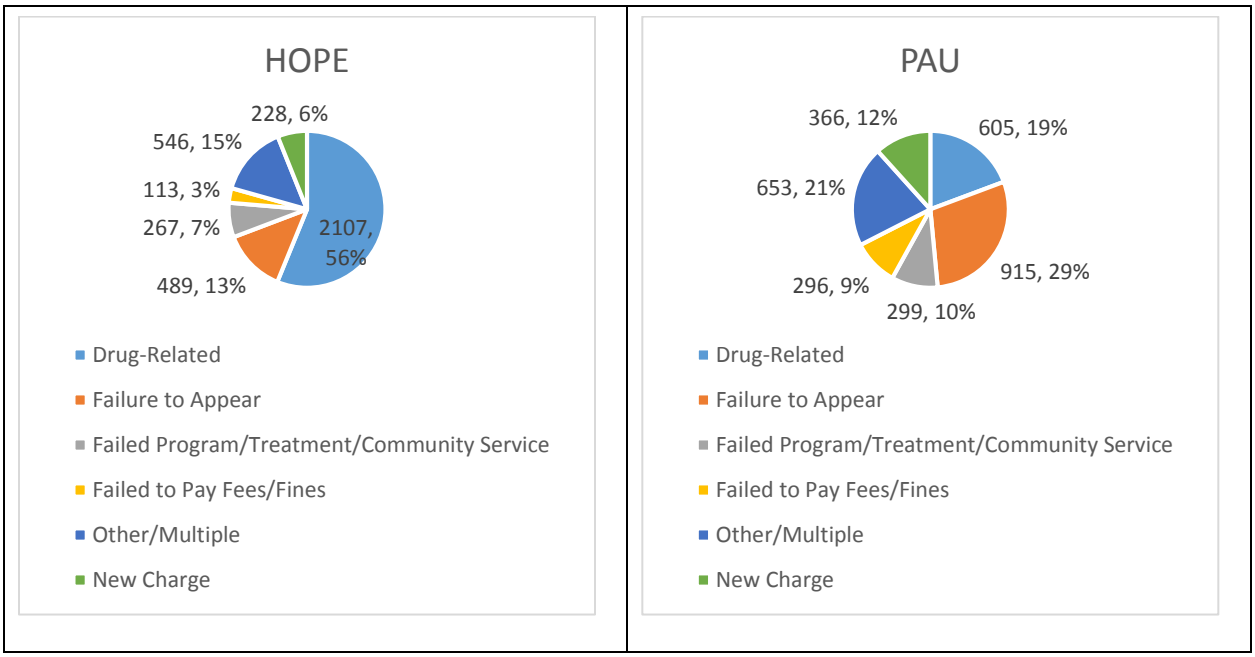
Note: Minimum counts were 0 for both groups for all violation types.
†HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.

HOPE probationers had fewer violations for failing to meet with their probation officers but more violations, on average, for failing to appear in court. Thus, HOPE seems to improve compliance with visits and the failures to appear in court may have been due to the HOPE requirement that every violation be followed by a violation hearing that was then missed by the probationer. There was no difference in the average number of violations for failing to complete programs or treatment and a significant but relatively small difference in the average number of violations for failing to complete community service requirements. ***HOPE probationers received half as many violations for failing to pay fees and fines.*** They received somewhat more violations for other types of violations and somewhat less for multiple causes on one violation. ***HOPE probationers had fewer violations for new charges—an average of 0.3 compared with nearly 0.5 for PAU.***

A final look at the nature of violations by group is shown in ***Exhibit 5-31***. Here, we have collapsed the drug-related charges (abstain from use, missed test), the failure-to-appear (probation officer visit,

court), the failure to complete program/treatment and community service, and the other and multiple counts to see the differences in distributions of violation types for the two study groups. As can be seen, ***more than half of the HOPE probationers' violations were drug-related, compared to less than 20% of the PAU probationers' violations.*** Again, the focus on frequent, random drug testing for the HOPE participants suggests that this finding is not surprising. ***The most common reason for a violation for the PAU group was for failure to appear either for a probationer officer visit or for court***—29% of PAU violations were for failure to appear compared with 13% of those for the HOPE group. Failure to pay fees and fines accounted for only 3% of the HOPE violations and 9% of the PAU violations. Violations for new charges were about 6% of the HOPE violations and 12% of the PAU violations.

Exhibit 5-31. Distribution of probation violations by group



Note: 743 HOPE probationers had a total of 3,770 violations; 761 PAU probationers had a total of 3,134.

There was variation in violation rates across the sites. Exhibit 5-32 shows the prevalence by violation type by site. ***Overall, there were fewer individuals cited in Massachusetts than in the other sites for either any or a specific type of violation.*** The proportion of individuals who were cited for any violation ranged from 0.74 in Massachusetts to 0.95 in Texas ($F = 27.19$; $p < 0.0001$). ***Drug-related violations were highest in Texas and Oregon.*** About 60% were cited for failing to abstain from drug use compared to 42% in Arkansas and 25% in Massachusetts ($F = 45.16$; $p < 0.0001$), while 45% of Texas probationers missed a drug test compared with 39% in Oregon, 29% in Arkansas and 20% in Massachusetts ($F = 21.14$; $p < 0.0001$).

Failing to appear for a probation officer visit was also most common in Texas and Oregon—49% and 45% compared with 30% in Arkansas and 23% in Massachusetts ($F = 26.13$; $p < 0.0001$). ***Failing to appear for court was less common overall***, but still most prevalent in Texas (21%) and Oregon (12%) compared to the other two sites ($F = 18.44$; $p < 0.0001$). Probationers in Texas and Oregon were also more likely to fail to complete programs or treatment ($F = 55.31$; $p < 0.0001$) and those in Texas were

most likely to fail to complete community service (F = 45.03; p < 0.0001).[37] ***Probationers in Texas and Arkansas were most likely to have received violations for failing to pay required fees or fines***—32% and 23% compared with 6% in Massachusetts and 1% in Oregon (F = 72.66; p < 0.0001). The sites were most similar with respect to citations for other types of violations—about a quarter of probationers in Arkansas, Oregon and Texas compared to 13% in Massachusetts (F = 7.06; p = 0.0001). ***The Texas probationers were least likely to have a violation for a new offense—9% compared with 16% in Arkansas, 33% in Massachusetts, and 40% in Oregon*** (F = 47.47; p < 0.0001).

Exhibit 5-32. Prevalence of violations by type and site (mean and standard deviation)

Type of Violation	AR	MA	OR	TX
Any violation=1* **	0.82 (0.38)	0.74 (0.44)	0.89 (0.31)	0.95 (0.21)
Fail to abstain from drug use=1* **	0.42 (0.49)	0.25 (0.44)	0.60 (0.49)	0.59 (0.49)
Missed drug test=1* **	0.29 (0.46)	0.20 (0.40)	0.39 (0.49)	0.45 (0.50)
Fail to appear PO visit=1* **	0.30 (0.46)	0.23 (0.42)	0.45 (0.50)	0.49 (0.50)
Fail to appear court=1* **	0.05 (0.22)	0.08 (0.27)	0.12 (0.32)	0.21 (0.41)
Fail to complete program/treatment=1* **	0.05 (0.22)	0.13 (0.33)	0.22 (0.41)	0.40 (0.49)
Fail to complete community service=1* **	0.04 (0.18)	0.01 (0.09)	0.02 (0.12)	0.17 (0.38)
Fail to pay fees/fines=1* **	0.23 (0.42)	0.06 (0.23)	0.01 (0.10)	0.32 (0.47)
Other type of violation=1* **	0.22 (0.41)	0.13 (0.33)	0.25 (0.44)	0.22 (0.41)
Multiple causes for single violation* **	0.27 (0.45)	0.24 (0.43)	0.05 (0.22)	0.53 (0.50)
New charge=1* **	0.16 (0.37)	0.33 (0.47)	0.40 (0.49)	0.09 (0.28)

***Sites differ at p < 0.001 for all types of violations.

We next examine site-level differences between the HOPE and PAU groups. The first rows ***in Exhibits 5-33*** and ***5-34*** show, for comparison purposes, the overall results that were reported previously. Results for HOPE and PAU in Arkansas, Massachusetts, Oregon, and Texas are then presented. Here, the differences among the eight groups (HOPE and PAU in four sites) are much more apparent. First, we see that ***the significant difference overall between the prevalence of any violation for HOPE and PAU is largely due to the difference in prevalence between the two groups in Arkansas***—where PAU probationers were significantly less likely to have experienced a violation than the HOPE probationers—72% versus 92%.

There are significant differences in the prevalence of violations for drug-related events between the HOPE and PAU groups in all four sites with HOPE probationers much more likely to have a drug-related violation. Again, this is a function of the nature of the HOPE program and an indication that the random drug testing and adherence to violations in response to any positive or missed tests were followed by the sites. There were also differences among the sites in the likelihood that a probationer would have a drug-related violation. Looking first at a violation for failing to abstain from drug use, we see that the percentage of HOPE probationers with at least one of this type of violation ranged from a

[37] It may be that probationers in these sites were more likely to be required to attend programs or treatment, resulting in them being more likely to fail to meet this condition. Similarly, Texas may have been more likely to assign community service.

low of 31% in Massachusetts to 62% in Arkansas, 71% in Oregon, and 76% in Texas. These contrasted with about 20% of PAU probationers in Arkansas and Massachusetts, 42% in Texas, and 50% in Oregon.

Exhibit 5-33. Prevalence of violations by type, site and group (mean and standard deviation)

Site	Group	Any	Fail to Abstain from Drugs	Missed Drug Test	Fail to Appear: PO Visit	Fail to Appear Court	N
All	HOPE	**0.89‡ (0.32)**	**0.60‡ (0.49)**	**0.58‡ (0.49)**	**0.30‡ (0.46)**	**0.18‡ (0.23)**	743
	PAU	**0.82‡ (0.39)**	**0.34‡ (0.47)**	**0.09‡ (0.29)**	**0.44‡ (0.50)**	**0.06‡ (0.23)**	761
AR	HOPE	**0.92‡ (0.28)**	**0.62‡ (0.49)**	**0.54‡ (0.50)**	0.27 (0.45)	0.05 (0.22)	179
	PAU	**0.72‡ (0.45)**	**0.21‡ (0.41)**	**0.02‡ (0.13)**	0.32 (0.47)	0.05 (0.22)	163
MA	HOPE	0.76 (0.43)	**0.31‡ (0.46)**	**0.39‡ (0.49)**	0.23 (0.42)	0.08 (0.27)	189
	PAU	0.72 (0.45)	**0.20‡ (0.40)**	**0.02‡ (0.16)**	0.24 (0.43)	0.07 (0.26)	203
OR	HOPE	0.91 (0.29)	**0.71‡ (0.45)**	**0.75‡ (0.43)**	**0.20‡ (0.40)**	**0.16† (0.37)**	190
	PAU	0.87 (0.33)	**0.50‡ (0.50)**	**0.04‡ (0.21)**	**0.68‡ (0.47)**	**0.08† (0.28)**	204
TX	HOPE	0.97 (0.18)	**0.76‡ (0.43)**	**0.63‡ (0.48)**	0.49 (0.50)	**0.42‡ (0.49)**	185
	PAU	0.94 (0.23)	**0.42‡ (0.50)**	**0.28‡ (0.45)**	0.50 (0.50)	**0.01‡ (0.10)**	191

†HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.

Differences were even greater for violations for missing a drug test—***75% of HOPE probationers in Oregon, 63% of those in Texas, 54% of those in Arkansas, and 39% of those in Massachusetts had at least one violation for missing a drug test.*** These percentages compare with 28% of those on PAU in Texas, 4% of those on PAU in Oregon, and 2% of those on PAU in Arkansas and Massachusetts who had at least one violation for missing a drug test. Two points are of note here. First, not everyone in HOPE in Massachusetts was subject to random drug testing because this condition had to have been imposed at sentencing (prior to assignment to HOPE) or following a significant event that could result in new conditions being imposed. Thus, the lower prevalence rates in Massachusetts for HOPE may simply reflect the fact that fewer individuals had required random tests. Second, except for Texas, very few individuals on PAU had a violation for a missed test likely reflecting the fact that there was little testing done in these sites under probation as usual.

There are fewer within-site differences between HOPE and PAU on the violations for failure to appear. ***The overall difference between HOPE and PAU on failure to appear for a probation officer visit (30% versus 44%) is driven by the difference between the two groups in Oregon, where 20% of HOPE probationers were cited for failing to appear for a visit compared with 68% of PAU.*** In Arkansas and Massachusetts, the percentages for both groups were comparable to the HOPE value in Oregon (i.e., in the 20% range) while values were much higher but similar for the two groups in Texas (i.e., about 50% of both groups). ***Failing to appear for court was significant overall and in Oregon and Texas; it was very low—in the 5% range—in Arkansas and Massachusetts and similar for the two groups.*** In Oregon, 16% of HOPE probationers and 8% of PAU probationers had at least one violation for failing to appear for court. These values were dwarfed by the percentages in Texas, where 42% of HOPE probations compared with 1% of PAU probationers had a violation for failing to appear in court. The high percentage in Texas for the HOPE group is likely an artifact of a local accommodation to the "rules of

HOPE." Specifically, in Texas, rather than issuing a warrant and arresting someone for violating a condition such as missing a visit or drug test, the HOPE probationer was given 24 hours to turn himself or herself in and to report to court. Thus, providing a higher likelihood that a court date would be set and missed.

Exhibit 5-34. Prevalence of violations by type, site and group, continued (mean and standard deviation)

Site	Group	Fail to Complete Program/ Treatment	Fail to Complete Community Service	Fail to Pay Fees/Fines	Other	Multiple	New Charge	N
All	HOPE	0.22 (0.41)	0.05 (0.21)	**0.11‡ (0.32)**	**0.23‡ (0.42)**	**0.24† (0.43)**	**0.22‡ (0.41)**	743
	PAU	0.19 (0.39)	0.06 (0.25)	**0.18‡ (0.39)**	**0.17‡ (0.38)**	**0.30† (0.46)**	**0.28‡ (0.45)**	761
AR	HOPE	0.05 (0.22)	0.05 (0.22)	0.22 (0.41)	**0.32‡ (0.47)**	0.24 (0.43)	**0.09‡ (0.29)**	179
	PAU	0.06 (0.23)	0.02 (0.13)	0.23 (0.42)	**0.10‡ (0.30)**	0.31 (0.46)	**0.31‡ (0.62)**	163
MA	HOPE	0.14 (0.35)	0.01 (0.07)	0**.03† (0.16)**	0.14 (0.35)	0.25 (0.44)	0.34 (0.47)	189
	PAU	0.11 (0.32)	0.01 (0.10)	**0.08† (0.28**)	0.11 (0.32)	0.23 (0.42)	0.33 (0.47)	203
OR	HOPE	0.24 (0.43)	0.01 (0.07)	**0.00† (0.00**)	0.23 (0.42)	**0.00‡ (0.00)**	**0.33‡ (0.47)**	190
	PAU	0.20 (0.40)	0.02 (0.16)	**0.02† (0.14)**	0.28 (0.45)	**0.10‡ (0.30)**	**0.47‡ (0.50)**	204
TX	HOPE	0.42 (0.50)	0.14 (0.34)	**0.22‡ (0.41)**	0.24 (0.43)	0.49 (0.50)	0.10 (0.30)	185
	PAU	0.37 (0.48)	0.20 (0.40)	**0.42‡ (0.50)**	0.19 (0.40)	0.58 (0.49)	0.07 (0.25)	191

†HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.

Overall, about 20% of all study participants were cited for failing to complete a program or treatment. There was tremendous variation across the sites, however, ranging from about 5% in Arkansas to 40% in Texas. There were no differences between groups in any of the sites and it is likely that the across-site variability is reflective of differences in requirements in those sites. For example, drug treatment was much more likely to be stipulated in Texas than in the other sites. There were also no differences in violations for failing to complete required community service—the percentages were very low everywhere with Texas, again, having the highest violation rate—14% for HOPE probationers and 20% for PAU probationers.

HOPE probationers were much less likely to be cited for failing to pay fees and fines than PAU probationers. This was true overall (11% versus 18%) and in three of the four sites (Massachusetts, Oregon, and Texas). Violation rates were highest in Texas—with those on PAU twice as likely as those on HOPE to have violations for failing to pay their fees and fines (22% versus 44%).

Individuals were also cited for multiple violations on the same day and for other events not previously specified. ***Overall, HOPE participants were more likely to have violations for 'other' actions while PAU probationers were more likely to have violations with multiple charges.*** These overall differences between the groups were driven in both cases by differences in one site. HOPE probationers were much more likely than PAU probationers to have violations for other charges (32% versus 10%) and PAU probationers were much more likely than HOPE probationers to have multiple charges (10% versus 0%).

HOPE probationers were much less likely to have a violation because of a new charge—22% versus 28%, overall. Significant differences between the two groups were observed in Arkansas (9% versus 31%) and Oregon (33% versus 47%), with no differences between groups in Massachusetts (about one-third of both groups) and Texas (10% or less).

5.3. Impact of Sanctions on Violations

One premise of HOPE is that swift and certain responses to violations will reduce future violations. Our data allow us to examine violations and the effect of sanctions on subsequent violations. The analyses took the form of sequential survival models (or "gap analyses")—each model conditioned on a positive outcome in the previous model, e.g., model 1 is time to first violation, model 2 is time to second violation conditioned on a first violation. (Observations were censored by the end of the study period, death, and incarcerated in prison; observations were not censored on jail days.) The models include indicators for site (Massachusetts, Oregon, and Texas; Arkansas was the reference category), revoked, and sanction other than jail days; as well as the number of jail days (capped at 48 days; 95th percentile) and the number of residential treatment days.

Results for the first five models are shown in ***Exhibit 5-35***. The first model suggests that time to first violation is longer for Massachusetts probationers compared with those in Arkansas, consistent with earlier discussion. ***The other models show that longer jail sanctions are associated with longer average times to subsequent violation*** (Gap 4 results are not statistically significant.). ***Treatment days are associated with shorter times to the next violation***—perhaps signifying more serious drug issues or more opportunity to violate by not complying with treatment requirements. ***Revocation is associated with a longer time to next violation***—likely an artifact of being removed from the community for the succeeding period. ***Having a sanction other than jail is also associated with a longer time to the next violation***, although most sanctions were jail.[38] Finally, we see little variability in the time between violations across the sites. Oregon has a longer mean time between the first and second violations and a shorter mean time between the fourth and fifth violations compared to Arkansas.

SWIFT AND CERTAIN SANCTIONS

HOPE is premised on the effectiveness of swift and certain sanctions to lead to future deterrence. Longer jail stays in response to violations were associated with longer times to the next violation.

The results discussed for Models 2-5 largely hold for subsequent models through Gap 10 (results not shown). We see positive significant effects of jail days and negative significant effects of treatment days on the times between violations.

[38] For example, the HOPE probationers received a total of 3,550 sanctions of which 2,920 (82.25%) were jail; in addition, 189 (5.32%) were revocations and 151 (4.25%) were treatment.

Exhibit 5-35. Results from lognormal survival models of time between violations

Variable	Coefficient (Standard Error)				
	Gap 1	Gap 2	Gap 3	Gap 4	Gap 5
Intercept	**6.53*** (0.25)**	**3.56*** (0.17)**	**3.03*** (0.19)**	**3.69*** (0.24)**	**3.56*** (0.27)**
Jail days		**0.02* (0.01)**	**0.03*** (0.01)**	0.01 (0.01)	**0.04*** (0.01)**
Treatment days		**-0.02*** (0.00)**	**-0.01*** (0.00)**	**-0.02*** (0.00)**	**-0.02*** (0.00)**
Revoked		**2.70*** (0.66)**	**3.91*** (0.63)**	**3.68*** (0.54)**	**4.50*** (0.53)**
Other Sanction		**1.37*** (0.32)**	**0.65* (0.33)**	0.63 (0.38)	0.24 (0.42)
MA (AR ref)	**1.06** (0.34)**	0.10 (0.25)	0.14 (0.24)	-0.38 (0.31)	-0.12 (0.34)
OR (AR ref)	0.05 (0.33)	**0.48* (0.23)**	-0.24 (0.24)	-0.38 (0.29)	**-1.36*** (0.30)**
TX (AR ref)	-0.22 (0.33)	0.33 (0.23)	-0.12 (0.25)	-0.19 (0.29)	-0.18 (0.32)
N	1504	663	575	501	406

*** p < 0.001, ** p < 0.01, * p < 0.05

Note: Gap 1 is the time between study intake and first violation, Gap 2 is the time between first and second violations, conditioned on having a first violation, etc.

5.4. Impact of HOPE on Jail and Prison Days

HOPE probation is premised on the use of short jail stays to sanction noncompliance in an expectation that these sanctions will deter future violations and criminal behavior. The cost of these additional jail days was expected to be more than offset by a reduction in prison days following revocations and incarcerations for new crimes. In the previous section, we saw that jail days were positively related to time between violations—i.e., that longer jail stays resulted in longer times to a new violation. This section looks at all jail stays and all prison stays not just stays that were the result of violations or revocations.

JAIL AND PRISON

Over both groups, the sites differed in the numbers of jail (1.4 to 4.0) and prison stays (0.13 to 0.38) and the total amounts of time individuals spent in jail (29 to 74 days) and prison (39 to 72 days).

Overall, including both groups, the sites differed in the average total number of jail days served by individuals—ranging from 22.8 days (sd = 26.5) in Texas, to 28.6 days (sd = 59.2) in Arkansas, 33.4 days (sd = 57.9) in Arkansas, and 73.9 days (sd = 95.6) in Oregon.[39] (Data not shown.) Although most individuals spent relatively few days in jail, there were outliers that influenced the means—including stays of more than 1 year in Arkansas (574 days), Massachusetts (469 days), and Oregon (715 days); the maximum stay in Texas was 138 days. The sites also differed in the average total number of jail stays—ranging from 1.4 (sd = 2.1) in Massachusetts to 1.8 (sd = 1.8) in Arkansas, 3.1 (sd = 3.5) in Texas, and 4.0 (sd = 3.9) in Oregon.[40]

[39] F = 49.41, p < 0.0001.

[40] F = 60.09, p < 0.0001.

Similar differences obtain for prison stays[41]. Total average prison days ranged from a low of 38.7 days (sd = 102.3) in Massachusetts to 45.3 days (sd = 145.9) in Oregon, 71.5 days (sd = 162.2) in Arkansas, and 71.6 days (sd = 110.9) in Texas.[42] The sites also differed in the average total number of prison stays—ranging from 0.13 (sd = 0.34) in Oregon to 0.21 (sd = 0.41) in Massachusetts, 0.23 (sd = 0.42) in Arkansas, and 0.38 (sd = 0.49) in Texas.[43]

Exhibit 5-36 compares the total number of jail stays and total jail days served by the HOPE and PAU groups, by site and overall. There were significant differences in total number of jail stays overall and in all sites. The maximum number of stays was 22 for HOPE probationers and 18 for PAU probationers (data not shown). Maximum numbers of stays for HOPE and PAU were 8 and 6 in Arkansas, 14 and 4 in Massachusetts, 22 and 18 in Oregon, and 20 and 8 in Texas. ***HOPE probationers also spent more days in jail—an average of 47.1 days compared with 33.3 for PAU probationers—as would have been expected if HOPE were implemented with fidelity. At the site level, however, the difference in jail days was statistically different in only two sites—Oregon and Texas—and was not different in Arkansas and Massachusetts.*** The average number of days was shorter, although not significantly shorter, in Arkansas, likely because of the higher revocation rate for HOPE probationers in Arkansas (as reported earlier).

Exhibit 5-36. Total numbers of jail stays and jail days by site and group

Site	Total Jail Stays			Total Jail Days		
	HOPE mean (sd)	PAU mean (sd)	t statistic	HOPE mean (sd)	PAU mean (sd)	t statistic
AR	**2.7*** (1.9)**	**0.8 (1.2)**	**-10.79**	31.1 (32.5)	36.0 (76.6)	0.76
MA	**2.2*** (2.7)**	**0.6 (0.9)**	**-7.949**	31.2 (58.5)	26.2 (59.9)	-0.833
OR	**5.0*** (4.0)**	**3.0 (3.5)**	**-5.072**	**89.6** (109.4)**	**59.3 (78.3)**	**-3.135**
TX	**5.2*** (3.8)**	**1.0 (1.2)**	**-14.371**	**35.4*** (30.3)**	**10.6 (14.0)**	**-10.134**
Overall	**3.8*** (3.5)**	**1.4 (2.3)**	**-15.533**	**47.1*** (70.8)**	**33.3 (65.0)**	**-3.955**

*** p < 0.001, ** p < 0.01

Exhibit 5-37 compares the total number of prison stays and total prison days served by the HOPE and PAU groups, by site and overall. There were significant differences in the total number of prison stays overall and in some sites. The maximum number of stays was 3 for HOPE probationers and 2 for PAU probationers (t = -2.094; data not shown). Maximum numbers of stays for HOPE and PAU were 3 and 2 in Arkansas, 1 and 2 in Massachusetts, 1 and 1 in Oregon, and 1 and 1 in Texas. ***Mean number of prison stays was significantly higher for HOPE overall—a result driven by significant differences in Arkansas and Oregon.*** (HOPE stays were slightly lower in Massachusetts and Texas, but the differences were not statistically significant.) ***The total number of prison days was also higher overall for HOPE (67.0 versus 45.5)—a result that was driven primarily by a large difference between HOPE and PAU in Arkansas (104.8 versus 34.9 days).***

[41] The length of a prison stay was assumed to be equal to sentence length.
[42] F = 6.47, p = 0.0002.
[43] F = 22.43, p < 0.0001.

Exhibit 5-37. Total prison stays and prison days served by site and group

Site	Total Prison Stays			Total Prison Days		
	HOPE mean (sd)	PAU mean (sd)	t statistic	HOPE mean (sd)	PAU mean (sd)	t statistic
AR	**0.37*** (0.58)**	**0.14 (0.38)**	**-4.423**	**104.8*** (190.7)**	**34.9 (113.3)**	**-4.162**
MA	0.19 (0.39)	0.23 (0.43)	0.978	40.3 (106.8)	37.2 (98.3)	-0.294
OR	**0.17* (0.38)**	**0.09 (0.29)**	**-2.213**	56.4 (165.6)	34.9 (124.3)	-1.454
TX	0.35 (0.48)	0.41 (0.49)	1.241	68.5 (112.1)	74.6 (110.0)	0.528
Overall	**0.27* (0.47)**	**0.22 (0.42)**	**-2.094**	**67.0** (149.2)**	**45.5 (112.9)**	**-3.147**

*** p < 0.001, ** p < 0.01, *p < 0.05

We also looked at average length of jail and prison stays—***these are the averages across stays not the average length of stay for individuals.*** There were 3,931 jail stays and 368 prison stays. Sites differed in the average length of stays. For jail stays, the mean ranged from 7.4 (sd = 7.6) in Texas to 18.7 (sd = 26.2) in Oregon, 19.3 (sd = 37.9) in Arkansas, and 23.1 (sd = 50.7) in Massachusetts.[44] Prison stays ranged from 182.7 (sd = 150.7; N = 83) in Massachusetts to 187.0 (sd = 102.7; N = 144) in Texas, 271.7 (sd = 187.7; N = 90) in Arkansas, and 349.7 (sd = 246.5; N = 51) in Oregon[45].

Exhibit 5-38 shows the average lengths of jail and prison stays by site and overall. Again, these are averages across stays not the average length of stay for individuals. HOPE probationers had 2,838 jail stays compared to 1,093 for PAU probationers. ***Overall and in three sites (Arkansas, Massachusetts, and Texas), the average lengths of jail stays were significantly shorter for HOPE than PAU; there was no difference in length of stay in Oregon.*** (Interestingly, Oregon PAU also had relatively more jail stays compared with the other sites—with about two-thirds the number of stays that Oregon HOPE probationers had, while PAU had only about one-quarter or so the number of stays that HOPE had in the other three sites. This may be because probation officers in Oregon can send probationers to jail for short stays on their own authority—without needing a judge's order.) ***Overall, prison sentences for HOPE were also longer—HOPE probationers served an average of 248.9 days in prison compared to 206 days for PAU46. This overall difference was driven by longer lengths of stay in three sites for the HOPE probationers, although these differences were not significantly different.***

LENGTH OF STAY

Average jail stays were shorter (overall and in three sites) and prison stays were longer (overall) for HOPE probationers than PAU probationers.

[44] F = 49.52, p < 0.0001.
[45] F = 17.20, p < 0.0001.
[46] Prison stay was equal to sentence length.

Exhibit 5-38. Average length of jail stays and prison stays by site and group

Site	Average Jail Days					Average Prison Days				
	HOPE		PAU		t	HOPE		PAU		t
	N	mean (sd)	N	mean (sd)		N	mean (sd)	N	mean (sd)	
AR	**492**	**11.6*** (15.2)**	**137**	**46.6 (69.5)**	**5.854**	67	280.0 (188.0)	23	247.6 (188.9)	-0.712
MA	**442**	**16.9*** (43.8)**	**135**	**43.1 (64.9)**	**4.402**	36	211.4 (155.2)	47	160.7 (144.9)	-1.530
OR	941	18.1 (26.0)	618	19.7 (26.5)	1.197	32	335.1 (266.3)	19	374.3 (200.4)	0.555
TX	**963**	**6.8*** (7.0)**	**203**	**10.1 (9.5)**	**4.690**	65	195.1 (105.1)	79	180.3 (100.8)	-0.857
All	**2838**	**13.0*** (24.6)**	**1093**	**24.2 (41.1)**	**8.473**	**200**	**248.9* (182.4)**	**168**	**206.0 (154.6)**	**-2.439**

Note: Averages are across jail and prison stays, not averages of total stays by individuals.
***p < 0.001, ** p < 0.01, * p < 0.05

5.5. Impact of HOPE on Drug Use

We have two sources of drug use data—administrative data reflecting the results of tests by the supervising agencies and oral swab drug test results conducted in conjunction with some follow-up interviews.[47] ***The administrative data do not provide a valid comparison of percent positive tests because only the HOPE probationers were subject to random testing*** and control group testing was more likely to be for cause (e.g., when the probation officer suspected that the individual was using) or scheduled. We can examine responses to positive tests within the HOPE-only group to assess the impact of sanctions on future test results. ***The oral swab tests conducted in conjunction with follow-up interviews were independent and do provide insight into comparative use.***

HOPE DRUG TESTING

The HOPE protocol recommended random twice weekly urinalysis drug testing for the first two months, followed by random weekly testing in the absence of any positive tests during the initial testing. The testing schedule would reset to more frequent testing in the event of a positive test.

Before presenting results, we first need to discuss the Massachusetts drug-test data. As described in the Methods section, electronic administrative drug-testing data were obtained from agencies in Arkansas, Oregon, and Texas, providing a consistent source of information for the HOPE and PAU groups. These electronic data were not available for Massachusetts. Instead, District Court case information was abstracted from pdf probation case summaries (generated by the Massachusetts Courts case management system; case summaries are input by the District Court probation officers) and the Superior Court information came from scanned probation officer "chrono" reports (case notes).

In some cases, these written records only included officer notes that the individual was complying with testing with no indication as to how many tests were conducted. We assume that positive test results are accurately reported (e.g., "So and so tested positive on <date>."), but we do not have

[47] Initially, oral swab tests were conducted during random weeks; towards the end of the follow-up period, tests were offered to all respondents who were not incarcerated at the time of the interview.

detailed information on negative tests (i.e., how many tests 'so and so' may have successfully passed). We do have drug test results from the fidelity data that were collected by HOPE project coordinators—but these data are only available for the HOPE probationers not the PAU probationers. ***There is a substantial (and meaningful) difference between the drug test data we could extract from the written records and the HOPE results from the fidelity data for our Massachusetts study subjects.*** Specifically, the administrative data provided information on 2,732 drug tests for HOPE probationers in Massachusetts while the fidelity data shows 5,515 tests for these individuals—so we could successfully identify and extract from the provided records specific information only on about 50% of the tests conducted on the HOPE probationers from the probationer officer files. The extracted data also showed 589 tests of PAU probationers—which is surely also an undercount but we cannot know whether this is also about 50% of all tests or some other fraction.[48]

In developing this section, we considered how to report administrative drug test findings for Massachusetts, faced with the following options:

1. Do not include Massachusetts in the analysis of administrative drug testing data;
2. Use the extracted data for both groups, knowingly undercounting the numbers of tests but assume that the undercount was similar for the two groups; or
3. Use the fidelity drug test data for HOPE and the extracted data for PAU, knowingly undercounting PAU but without knowing by how much.

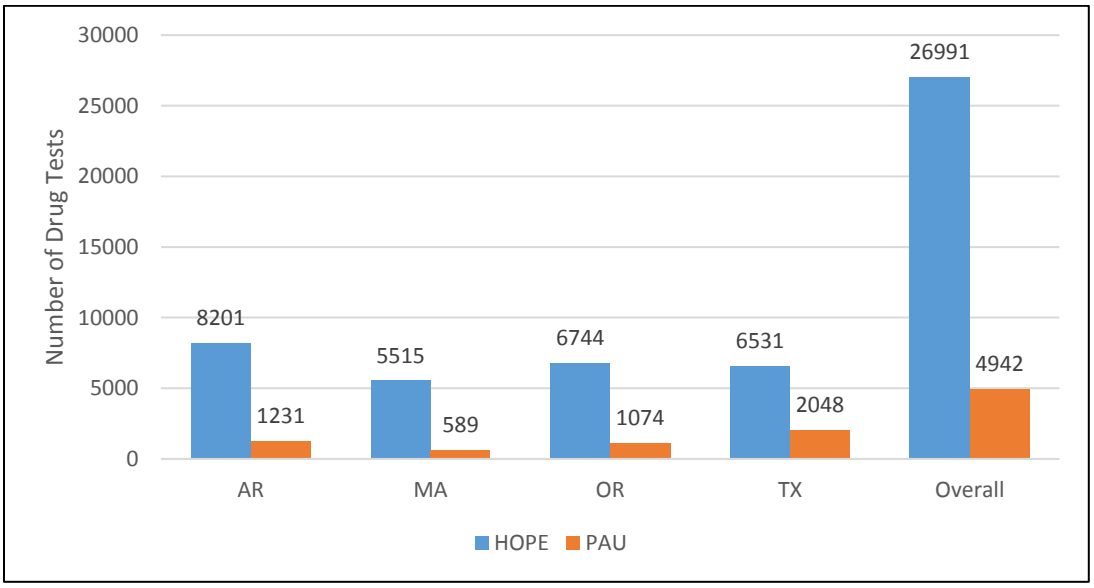
We decided the third option was best as it makes the most use of the information we have. As the test results between HOPE and PAU are not directly comparable anyway (because there was not random testing for PAU probationers), using the fidelity data allows a better comparison across the four sites for the use of drug testing among those on HOPE probation.

Administrative Drug Test Results: HOPE and PAU

Of the 1,504 study subjects, 1,275 (85%) were drug tested a total of 31,933 times per local administrative records—84.5% of these tests (26,991) were administered to 685 HOPE probationers and 15.5% (4,942) were administered to 590 PAU participants. ***HOPE participants were tested an average of 39.4 times and PAU probationers were tested an average of 8.4 times. Exhibit 5-39*** shows the number of tests performed by group and site. As can be seen, fewer tests were conducted in Massachusetts than in the other three sites—only about 19% of all tests were conducted in Massachusetts compared to 30%, 24%, and 27% in Arkansas, Oregon, and Texas. (Reminder that the drug test count for Massachusetts PAU probationers is likely an undercount of the total tests that were conducted.)

[48] We also checked the fidelity test counts against the administrative test counts for the HOPE participants in the other sites. These counts were very similar. Fidelity counts and administrative counts for Arkansas were 7,988 and 8,201, for Oregon were 6,744 and 6,744, and for Texas were 6,499 and 6,531.

Exhibit 5-39. Total number of drug tests administered to study participants by site and group

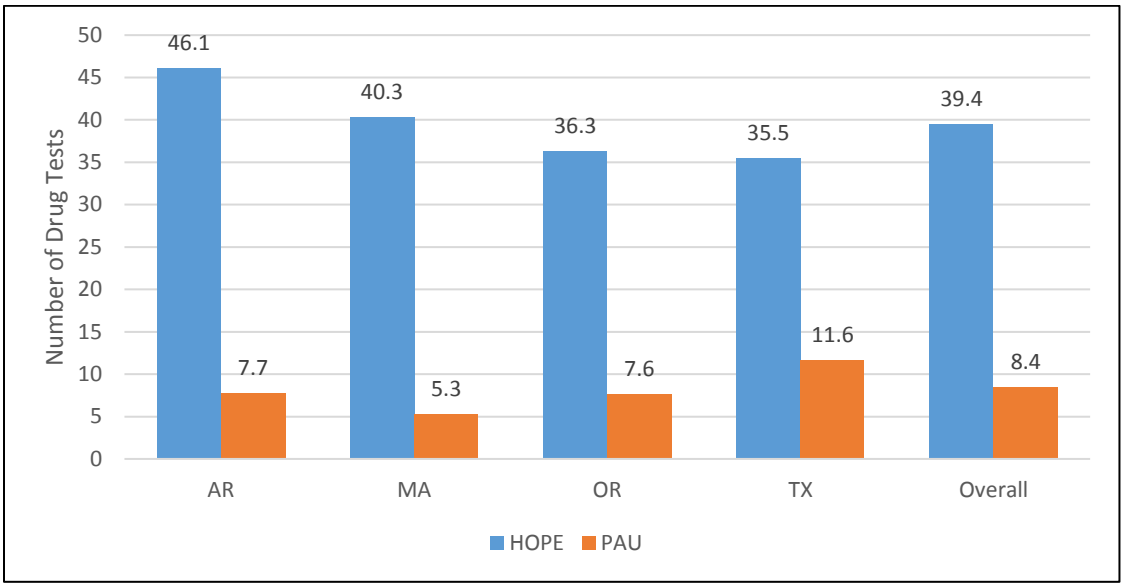


Note: HOPE and PAU data from Massachusetts were extracted from different sources and are not directly comparable.

Exhibit 5-40 shows the average number of tests per person by group and site. Among HOPE probationers, Arkansas had the highest average number of drug tests per probationer (46.1), and Texas had the lowest (35.5). Note that while the Massachusetts HOPE group experienced the lowest number of drug tests, the average number of tests per probationer (40.3) rivaled Arkansas since, compared to the other sites, fewer Massachusetts HOPE probationers were subjected to testing.[49] Testing for the PAU groups ranged from about 5 in Massachusetts to 12 in Texas. As noted previously, most of the PAU testing was not random testing but testing for cause or scheduled.

[49] About 73% of HOPE probationers in Massachusetts were subjected to drug testing, compared to 98% in Oregon and 99% in Arkansas and Texas.

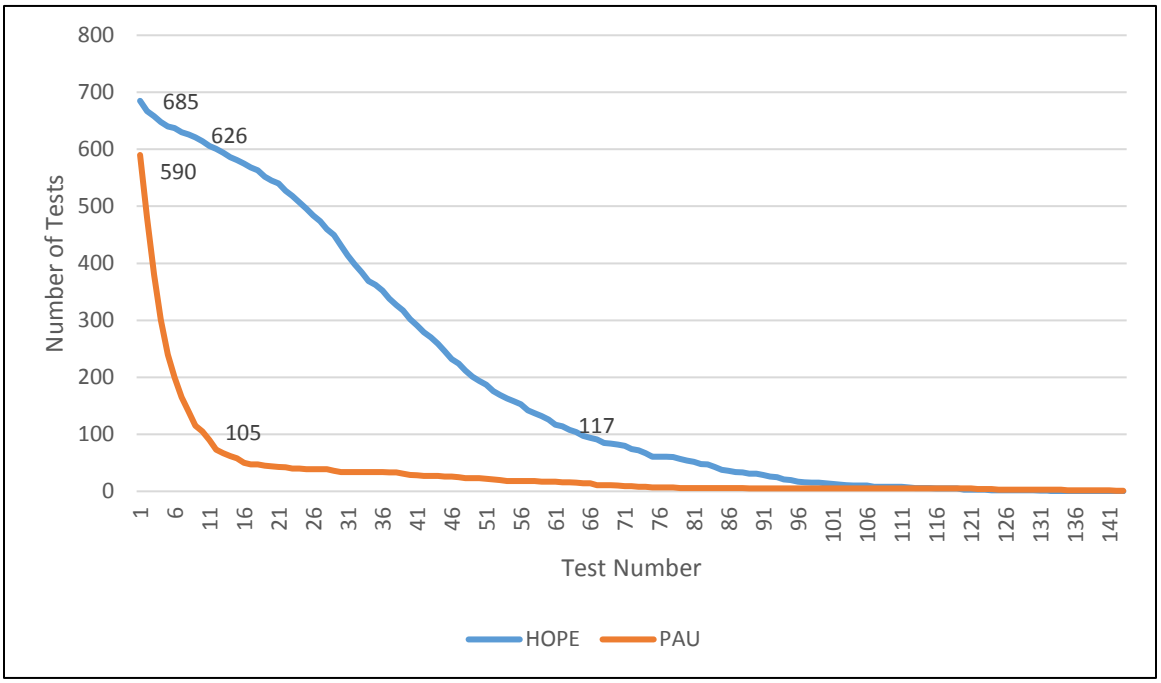
Exhibit 5-40. Average number of drug tests administered to study participants by site and group



Note: HOPE and PAU data from Massachusetts were extracted from different sources and are not directly comparable.

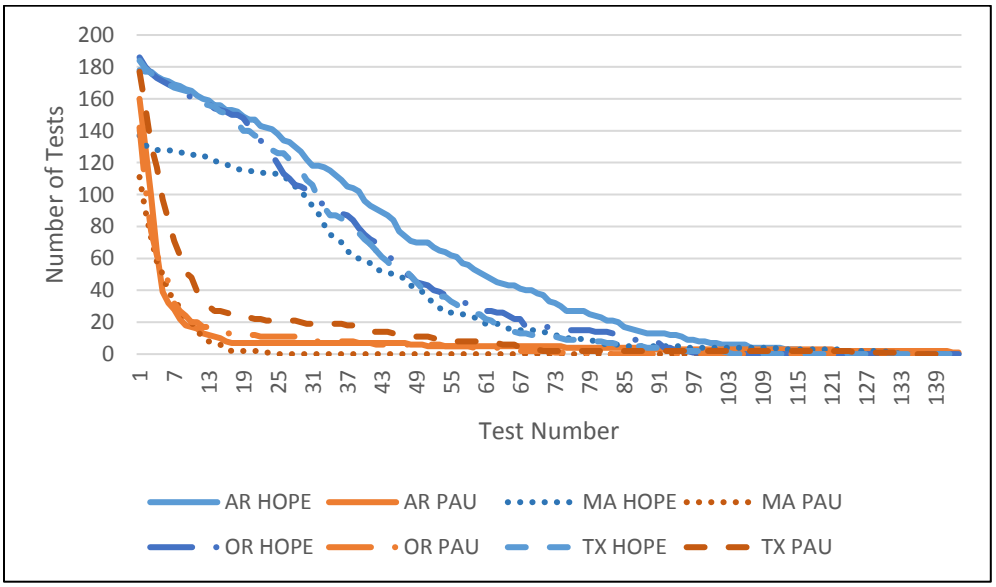
Exhibit 5-41 shows the number of individuals who were tested one, two, three, etc. times. As can be seen, not everyone in either group was tested even once—685 of 743 (92%) HOPE probationers were tested at least once, while only 590 of the 761 (78%) PAU probationers were tested at least once. Although all HOPE cases should have been tested, the discrepancy is primarily due to policy/legal considerations in Massachusetts that meant that some HOPE cases could not be required to participate in the random drug testing if the testing requirement was not imposed at sentencing (prior to assignment to HOPE probation). Specifically, only 137 of 189 HOPE probationers in Massachusetts were tested at least once—or 72%. ***Exhibit 5-41*** also shows that not only were HOPE probationers more likely to be tested—but, as expected, they were tested many more times than the PAU probationers. Indeed, only 105 of the PAU probationers had 10 or more drug tests, compared with 626 of the HOPE probationers. More than 100 of the HOPE probationers were tested 66 times (or more).

Exhibit 5-41. Drug testing of HOPE DFE study participants by group



The distributions of testing across the sites and groups are shown in ***Exhibit 5-42***. There is some variation, but the patterns are like the overall patterns. The smaller likelihood of any tests for HOPE probationers in Massachusetts is apparent on this graph as study groups were of similar size.

Exhibit 5-42. Drug testing of HOPE DFE study participants by group



Note: HOPE and PAU data from Massachusetts were extracted from different sources and are not directly comparable.

Exhibit 5-43 shows the percentage positive for any drug for the first 12 drug tests by site for the HOPE groups. (PAU percent positives are not comparable and are not shown.) The percentage testing positive is higher in Massachusetts than the other sites—at least after the first few tests. By the fifth test, the percent positive dropped to 10% or less in all sites except Massachusetts.

Exhibit 5-43. Percentage positive drug tests for HOPE probationers by site and test number

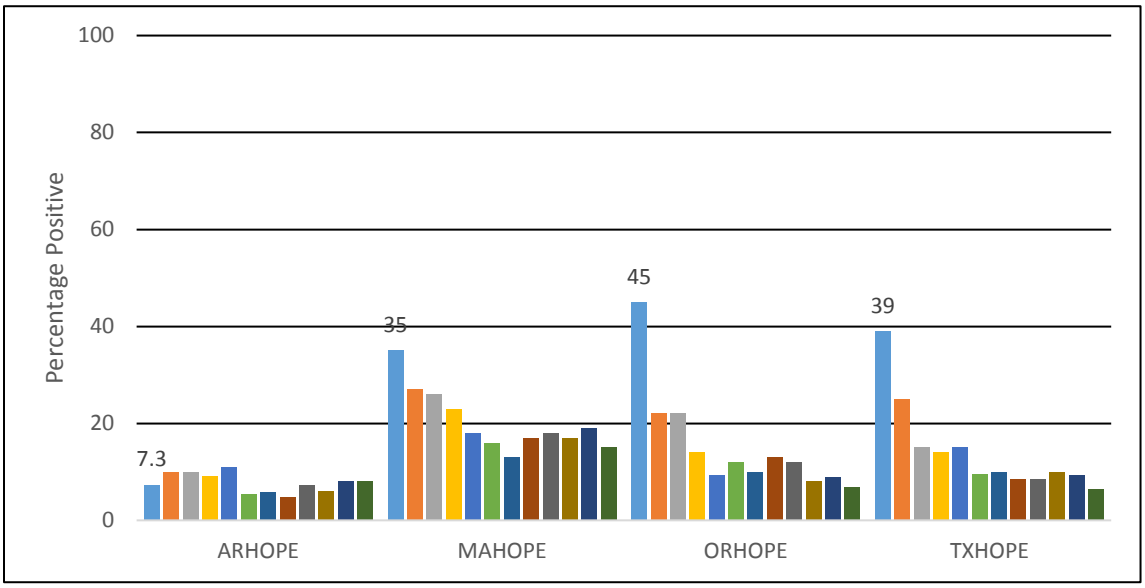
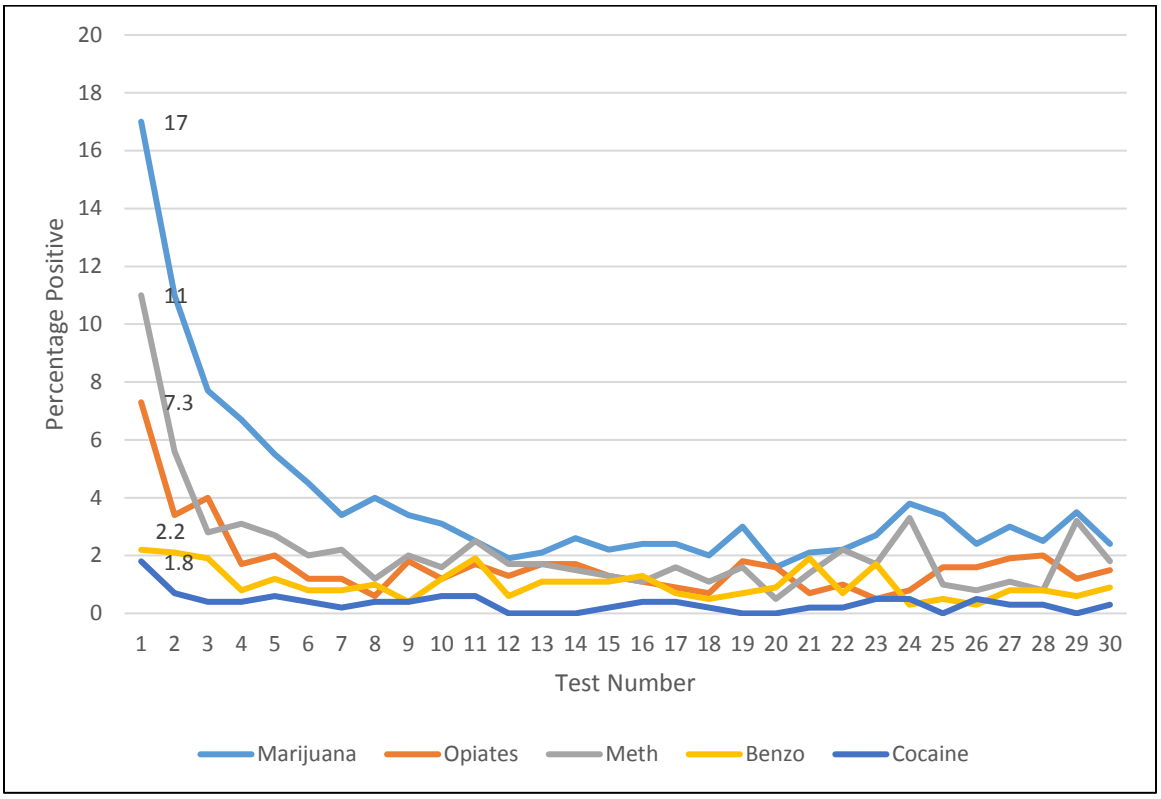


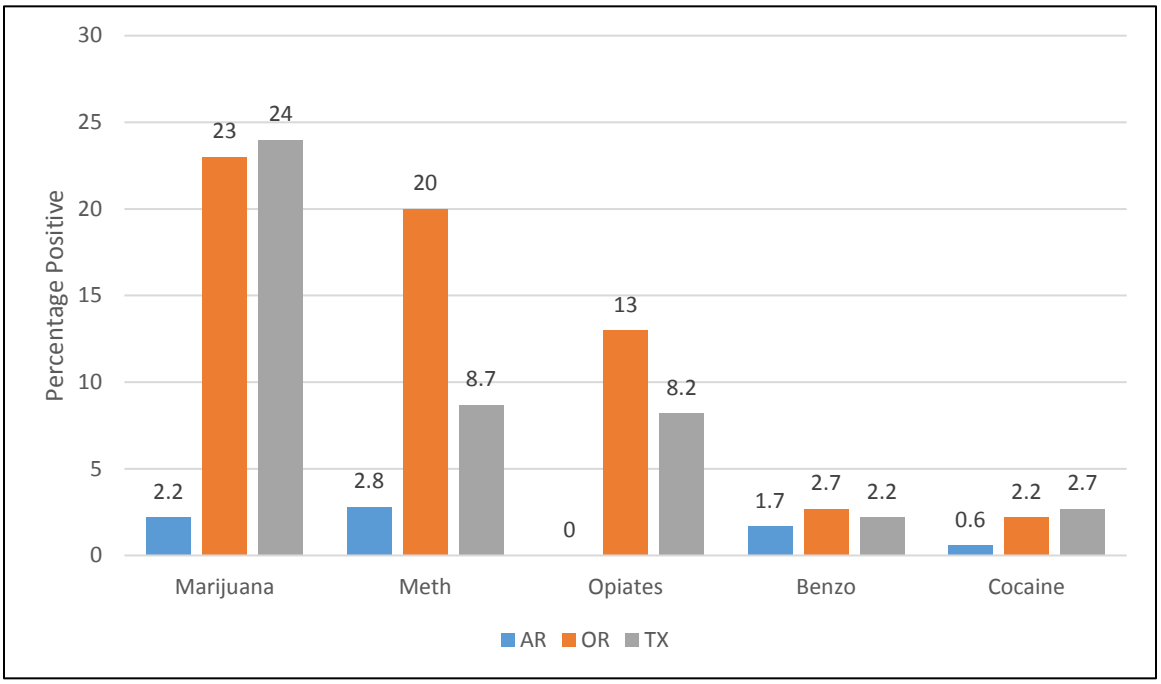
Exhibit 5-44 shows the percentages of positive tests for the HOPE groups by drug type for the first 30 tests. (Drug-specific test results were not available for Massachusetts.) Seventeen percent of the HOPE probationers tested positive for marijuana on the first drug test. This percentage declined rapidly and by the sixth test only 5% tested positive for this drug. The next most common positive test results were for methamphetamine (11% positive on the first test) and opiates (7.3% positive on the first test). Positive results were much less common for benzodiazepines (2.2%) and cocaine (1.8%). As shown in the exhibit, ***positive tests dropped rapidly for all drug types and remained low throughout the testing.***

Exhibit 5-44. Percentage positive drug tests for HOPE probationers (AR, OR, TX) by drug type and test number



Site-specific results for positives on the first drug test are shown in ***Exhibit 5-45***. Any positive test was rare in Arkansas***. Marijuana was the "drug of choice" in Oregon and Texas (23% and 24% testing positive), while methamphetamine was the second most prevalent in those two sites (20% and 8.7%), followed by opiates (13% and 8.2%).***

Exhibit 5-45. Percentage positive by drug type for HOPE probationers, first test



Administrative Drug Test Results and Responses to Sanctions: HOPE Only

Logistic regression models were estimated to assess the impact of sanctions on urinalysis results for the HOPE participants. Specifically, for individuals with a positive test, the results of the next test was regressed on the length of sanction associated with the previous positive test and the test number as a control. (Results are for the first positive and subsequent test only; total N = 468.) We also included site and a jail days*site interaction. Results (***Exhibit 5-46***) suggest that higher test numbers are associated with less risk of a consecutive positive test. ***More jail days are associated with a greater risk that the next test will be positive.*** There were no site-level effects and marginal interaction effects (p values 0f 0.06 or 0.07)—suggesting that the effect of time in jail is less for the other sites than for Arkansas.

Exhibit 5-46. Logistic model results for positive drug test results (HOPE only)

Variable	Estimate	Standard Error	Odds Ratio	Z statistic
Intercept	**-1.277*** **	**0.297**	0.279	**-4.304**
Test number	**-0.031***	**0.013**	0.969	**-2.476**
Jail days (t-1)	**0.080****	**0.029**	1.083	**2.727**
MA (AR ref)	0.056	0.436	1.058	0.128
OR (AR ref)	0.411	0.359	1.508	1.143
TX (AR ref)	-0.070	0.388	0.932	-0.182
Jail days*MA	-0.091	0.051	0.913	-1.793
Jail days*OR	-0.061	0.034	0.941	-1.816
Jail days*TX	-0.070	0.038	0.932	-1.844

*p < 0.05; **p < 0.01, *** p < 0.001.

RTI-Conducted Oral Swab Drug Test Results: HOPE and PAU

Oral swab drug testing was conducted in conjunction with some follow-up interviews (in the community, on a random schedule for most of the study period). As not everyone was interviewed and not everyone who was interviewed was asked to consent to the oral swab test, we had four possible results:

1. No interview
2. Interview, no test
3. Interview, negative test
4. Interview, positive test

Distributions at 6- and 12-month follow-up interviews are shown in ***Exhibits 5-47*** and ***5-48***. (Excluded are those incarcerated, who weren't asked to participate in the oral swab tests, and those who were deceased.) As previously discussed, the most common is no interview and response bias analyses suggested those who were interviewed were like those who were not. We focused our analyses on those who participated in the oral swab drug testing—which was offered on a random week basis during the early part of the study follow-up and for everyone in the community during the final months of follow-up data collection.

Exhibit 5-47. Interview/no interview and oral swab results at 6-month follow up

Site	Group	6 Month No Interview	6 Month Interview, No Test	6 Month Interview, Negative Test	6 Month Interview, Positive Test	Total
AR	HOPE	78	49	43	4	174
	PAU	106	25	17	9	157
MA	HOPE	96	60	19	9	184
	PAU	113	55	17	12	197
OR	HOPE	96	53	17	2	168
	PAU	128	46	11	1	186
TX	HOPE	86	36	14	2	138
	PAU	123	29	9	1	162
Total	HOPE	356	198	93	17	664
	PAU	470	155	54	23	702
Total		826	353	147	40	1366

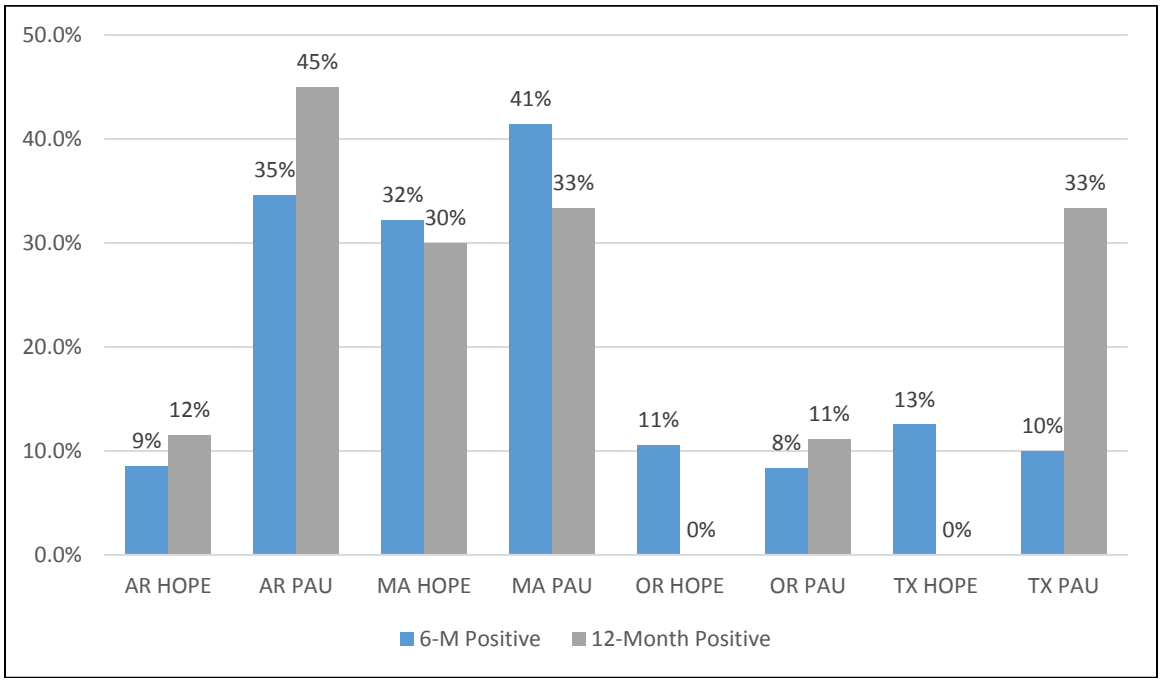
Overall, 15.5% of the HOPE probationers had a positive oral swab drug test compared to 29.9% of the PAU probationers at the 6-month interview. Similar results obtained at the 12-month interview, with 12.5% of the HOPE probationers testing positive compared to 31.2% of the PAU probationers.

Exhibit 5-48. Interview/no interview and oral swab results at 12-month follow up

Site	Group	12 Month No Interview	12 Month Interview, No Test	12 Month Interview, Negative Test	12 Month Interview, Positive Test	Total
AR	HOPE	56	52	23	3	134
	PAU	58	35	11	9	113
MA	HOPE	94	38	14	6	152
	PAU	109	26	22	11	168
OR	HOPE	108	54	6	0	168
	PAU	136	33	16	2	187
TX	HOPE	71	32	20	0	123
	PAU	123	16	4	2	145
Total	HOPE	329	176	63	9	577
	PAU	426	110	53	24	613
Total		826	755	286	116	33

Exhibit 5-49 shows the percent testing positive by site and group for the two waves. As can be seen the PAU probationers were more likely to test positive than the HOPE probationers in Arkansas and Massachusetts at both rounds of testing. Results were similar for the two groups at the 6-month interview in Oregon and Texas—and like the Arkansas HOPE results of about 10%. At 12 months, none of the Oregon or Texas HOPE swabs were positive and about 11% of the PAU samples in Oregon were positive. One-third of the PAU samples in Texas at 12 months tested positive although the overall numbers were small (2 of 6 samples).

Exhibit 5-49. Oral swab test results by site, group, and follow-up wave



Logistic regression results examining the impact of HOPE participation (and site) on drug test outcomes are shown in ***Exhibit 5-50***. Forty of the 185 tests had a positive result. As can be seen, confirming the bivariate results, participation in HOPE was associated with reduced odds of testing positive during both the 6- and 12-month follow-up interview[50]. Probationers in Massachusetts were more likely than those in Arkansas to test positive at 6 months; probationers in Oregon were less likely than those in Arkansas to test positive at 12 months.

Exhibit 5-50. Logistic regression results for oral swab drug test results

Variable	6 Month Post Baseline			12 Months Post Baseline		
	Estimate	Std. Error	Odds Ratio	Estimate	Std. Error	Odds Ratio
Intercept	-1.070	0.412	0.343	-0.453	0.449	0.635
HOPE	**-0.734**	**0.667**	**0.480**	**-1.185****	**0.761**	**0.306**
MA (AR reference)	**0.874***	**0.559**	**2.396**	0.086	0.582	1.090
OR (AR reference	-0.766	1.123	0.465	**-1.685***	**0.877**	**0.185**
TX (AR reference)	-0.565	1.132	0.568	-1.238	0.976	0.290

* p < 0.05; **p < 0.01

5.6. Impact of HOPE Probation on Attitudes

The T-ACASI mini-study allowed us to examine the impact of HOPE probation on attitudes. Attitudinal scales that were a part of the ACASI interviews were also measured during the telephone call-ins from those who agreed to participate. These attitudinal scales were Family Support, Un-cynical Attitudes, Treatment motivation, Probation Officer Attitude, Deterrence, Readiness for Change, Self-Efficacy, Identification with Criminal Others, and Attitudes toward the Law. We used bivariate logistic regressions to assess whether attitudinal, criminal history (age at first arrest, number of prior arrests, number of convictions, juvenile detention, number of times incarcerated), and demographic variables (sex, race, and age at baseline) were associated with the likelihood to consent to participating in T-ACASI interviews. All participants in Arkansas consented, resulting in zero variance and no models were fit. None of the background variables predicted consent in Massachusetts or Oregon. In Texas, participants with higher scores on the uncyncial attitudes scale were more likely to consent (i.e., those who were less cynical), suggesting that this is the only measure and site where mixed model results might have issues with response bias.[51] We tested whether there were differences between those who completed only the baseline ACASI scales and those who completed the baseline plus at least one T-ACASI scale (i.e., contrasting those who consented but never called against those who consented and participated in the mini-interview at least once). There were no significant results, indicating that consenting but not participating was unrelated to other study variables and suggesting that the assumptions about missingness made by our mixed models held in this sample.

[50] The t-statistic for the 6-month test was -1.953, p = 0.051.

[51] Analyses were not conducted for females in OR and TX, black participants in Massachusetts and for race in OR because there was no variance on these variables.

The frequency of participants completing each attitudinal scale from one (baseline) to nine times is shown in ***Exhibit 5-51***, where we see that responding tapers after the second interview—or after the first time that participants called in (the column labeled 3).

Exhibit 5-51. Counts of respondents completing attitudinal measure scales by site by group

Scale	Site	HOPE									PAU								
		1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
Attitudes Toward the Law	AR	17	5	3	2	2	1	1	0	0	12	2	1	0	0	0	0	0	0
	MA	56	20	10	5	4	3	0	0	0	61	23	18	14	9	5	3	1	0
	OR	46	14	11	5	3	0	0	0	0	42	14	9	5	2	0	0	0	0
	TX	27	4	4	2	2	0	0	0	0	21	5	2	2	2	0	0	0	0
Identification with Criminal Others	AR	17	5	5	3	2	2	0	0	0	12	1	1	0	0	0	0	0	0
	MA	56	14	11	8	3	3	2	0	0	61	23	15	14	9	7	3	2	0
	OR	46	14	7	7	3	1	1	1	0	42	11	7	3	1	1	0	0	0
	TX	27	4	1	1	1	0	0	0	0	21	5	3	3	2	1	0	0	0
Self-Efficacy & Readiness for Change	AR	17	6	5	2	2	1	1	0	0	12	1	1	0	0	0	0	0	0
	MA	56	16	10	7	4	1	1	1	0	61	21	17	13	10	7	2	1	0
	OR	46	14	9	4	3	1	1	1	0	42	14	6	3	2	2	1	0	0
	TX	27	4	2	1	1	1	1	0	0	21	5	5	2	1	0	0	0	0
Deterrence	AR	17	8	4	2	2	1	0	0	0	12	1	1	1	1	0	0	0	0
	MA	56	15	8	8	7	4	2	1	0	61	21	14	12	9	6	2	0	0
	OR	46	15	6	2	1	1	1	0	0	42	9	8	5	5	1	0	0	0
	TX	27	4	3	2	1	1	1	1	0	21	4	4	3	0	0	0	0	0
PO Attitude	AR	17	4	3	2	2	1	0	0	0	12	2	2	1	0	0	0	0	0
	MA	56	16	9	7	4	3	2	1	0	61	25	12	10	8	6	1	1	0
	OR	46	16	10	5	2	0	0	0	0	42	13	10	4	4	3	0	0	0
	TX	27	7	3	2	1	1	0	0	0	21	5	5	2	2	1	0	0	0
Treatment Motivation	AR	17	6	3	3	2	2	1	0	0	12	1	1	1	1	0	0	0	0
	MA	56	15	11	6	4	2	1	0	0	61	22	15	12	9	5	3	0	0
	OR	46	15	6	2	1	0	0	0	0	42	13	9	4	2	0	0	0	0
	TX	27	7	2	1	1	0	0	0	0	21	4	3	1	1	1	0	0	0
Uncynical Attitudes	AR	17	6	3	2	2	0	0	0	0	12	1	1	1	0	0	0	0	0
	MA	56	15	9	5	3	1	1	0	0	61	20	13	8	8	1	0	0	0
	OR	46	16	10	5	1	0	0	0	0	42	13	9	4	1	0	0	0	0
	TX	27	5	2	2	1	1	1	0	0	21	7	3	3	2	1	1	0	0
Family Support	AR	17	7	2	2	2	1	0	0	0	12	1	1	0	0	0	0	0	0
	MA	56	16	11	8	5	2	1	1	1	61	19	18	11	8	4	4	1	0
	OR	46	10	5	3	3	0	0	0	0	42	11	7	5	2	1	1	1	0
	TX	27	6	1	1	0	0	0	0	0	21	5	5	3	2	0	0	0	0

NOTE: First interview is the baseline interview so the counts indicate the numbers of individuals who consented.

Between-Site Differences in Attitudes

Exhibit 5-52 shows the means and standard deviations for the nine scales at baseline by site. (Scales were discussed previously in section 4.6; a copy of the T-ACASI instrument is included in Appendix E.) We compared the four sites' means on each scale from the baseline interviews using pairwise t-tests with a Bonferroni correction. Higher scores are better for all scales except the Identification with Criminal Others scale. Scores overall were on the "better" end, which may reflect some social desirability response bias on the part of respondents. As we are interested in trends over time this may be less important for these analyses than if we were interested in specific values.

Exhibit 5-52. Means and standard deviations of responses on each attitudinal scale at baseline by site

Scale (range)	Range		Arkansas		Mass.		Oregon		Texas	
	Min	Max	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Family support (14-126)	17	126	91.12	27.59	88.07	27.1	86.67	26.7	96.25	27
Uncynical attitudes (5-45)	5	45	35.93	9.32	33.61	9.56	35.27	9.15	34.83	8.87
Treatment motivation (10-90)	10	90	39.54	20.46	41.11	17.46	44.83	18.18	37.32	13.11
PO attitudes (6-54)	6	54	34.2	14.17	35.98	15.15	37.67	14.37	41.08	12.71
Deterrence (10-90)	26	90	61.5	13.32	**61.44**[1]	13.21	63.17	11.09	**68.03**[1]	6.83
Readiness for Change (4-36)	4	36	29.75	6.7	30.61	8	31.31	7.61	31.63	6.05
Self-efficacy (4-36)	6	36	26.28	6.61	26.58	6.98	28.1	6.01	28.7	5.88
CSS-M Identification with Criminal Others (0-12)	0	12	3.66	1.86	**4.18**[2]	2.21	**3.16**[2]	2.1	3.95	2.32
CSS-M Tolerance for Law Violations (0-20)	4	16	10.07	1.85	10.34	1.73	10.21	2.21	9.95	1.79

[1] Baseline deterrence was significantly higher in Texas than Massachusetts.
[2] Mean identification with criminal others was significantly higher in Massachusetts than Oregon.
Note: Pairwise comparisons were conducted using Bonferroni corrected t-tests. For all variables except identification with criminal others, higher scores are better.

The ***Family Support Scale*** included 14 items, including measures of "closeness" ("I feel close to my family."), negative interactions ("I fight a lot with my family members."), and supports ("I have someone in my family who would provide financial support."). The maximum value for this scale was 126 and mean scores for respondents across the four sites ranged from 87 (Clackamas County, Oregon) to 96 (Tarrant County, Texas)—with no significant between-site differences—suggesting that respondents reported on average having supportive family.

The ***Uncynical Attitudes Scale*** included five items, also scored 1 to 9 as described above, such as "Laws are made to be broken." These items were intended to measure respondents' views about the fairness of and their cynicism towards the legal system. Items were reverse coded so that a higher score suggested less cynicism and the maximum score was 45. Again, there was no differences among the sites, with mean scores ranging between about 34 and 36.

The ***Treatment Motivation Scale*** (a subscale of the TCUDS) included eight items, which were scored on the 1-to-9 scale and the average response of non-missing items multiplied by 10 to yield the score, providing a maximum value of 90. Unlike the previous scales, the average respondent was slightly more

likely to disagree with the various items, suggesting lower motivation for treatment. Average scores ranged from 37 to 45, with no significant differences between sites.

The ***Probation Officer Attitudes scale*** included six items, which were again scored on a 1-to-9 scale, with a maximum value of 54. This scale was intended to measure the respondent's impressions of their probation officer's treatment and included items such as "My probation officer helps me to succeed." Mean scores for participants in the four sites were above average and ranged from 34 to 41, with no significant differences between sites.

The ***Deterrence scale*** included 10 items, scored on a 1-to-9 scale, with a maximum value of 90. This scale was intended to measure the respondent's beliefs with respect to the consequences of not complying with the conditions of supervision including admitting illegal drug use, testing positive for illegal drugs, skipping drug tests, missing appointments, or failing to attend treatment. Questions included items related to the probation officer's responses and the judge's responses. Mean scores were above average and ranged from 34 to 41, with only one significant between-site difference—the mean score is significantly higher in Tarrant County than in Essex County, suggesting the Texas probationers were more likely to expect consequences from their probation officers or judges in response to a violation of supervision conditions.

The ***Readiness for Change scale*** included four items, scored on a 1-to-9 scale, with a maximum value of 36. This scale included measures such as "I am working to get my life straightened out." And "I am willing to accept help in dealing with staying straight." Mean scores were well above average ("neither disagree nor agree") and ranged from 30 to 32, with no significant differences between sites.

The ***Self-Efficacy scale*** also included four items, again scored on a 1-to-9 scale, with a maximum value of 36. This scale was intended to measure the respondent's beliefs with respect to the ability to affect what happens including "I have little control over the things that happen to me." and "My life has gone out of control." Some items were reverse scored so that higher scale scores suggested higher self-efficacy. Mean scores were above average and ranged from 26 to 29, with no significant differences between sites.

The final two scales are subscales from the CSS-M and were scored with "1" if the respondent disagreed with each item, "2" if the respondent agreed, and "3" if the respondent was unsure. The ***Identification with Criminal Others scale (CSS3)*** included six items, such as "People who have broken the law have the same sorts of ideas about life as me." Higher scores on this scale are "worse" since higher scores suggest more identification with criminal others. The maximum value on this scale was 12, and mean scores across the four sites were quite low, ranging from 3 to 4. ***There was one between-site difference with the mean scores significantly higher in Massachusetts than Oregon (4.18 versus 3.16).*** This difference may be due to differences in social desirability bias or possibly reflect a true population difference as the Massachusetts sample included individuals who had served substantial prison sentences prior to being placed on probation (from the Superior Court).

The ***Tolerance for Law Violations scale (CSS2)*** included 10 items, with a maximum value of 20. Questions included items such as "Most successful people broke the law to get ahead in life." and "It's okay to break the law as long as you don't get caught." Mean scores in each site were at the average of 10 ("neither disagree nor agree") suggesting tolerance for breaking the law, with no significant differences between sites.

Thus, overall, despite demographic and criminal history differences among the sites' HOPE-eligible populations, we see broad similarities in attitudes at baseline across the sites.

Within-Site Differences in Attitudes between HOPE and PAU Respondents

The next set of analyses assessed whether there were differences between the HOPE and PAU groups within each site at baseline. There were no significant differences between the groups, as would be expected if randomization to treatment worked as expected.

Changes over Time in Attitudes

Daily average trajectories and best linear fit lines for each scale by group are shown in ***Exhibit 5-53***. These charts allow us to examine whether there are changes in attitudes over time and whether any changes differ between HOPE and PAU probationers. For example, looking at the family support scale results, allows us to assess whether family support is increasing (or decreasing) over time. Each participant's response is an estimate of overall family support and, since different participants respond each day, there is wide day-to-day variation in average family support (the grey lines). However, ***on average, family support improved for the HOPE group, but was stagnant for the PAU group*** (the red lines). Whether this difference is significant is shown by the interaction term in the mixed models that are presented next. Averaging over the two groups would yield a family support trend line that increases, but not as steeply as shown for the HOPE group. Whether the increase in this averaged line is significant is shown by the main effect for time in the mixed models. Averaging the group difference across time would indicate that the HOPE group had better family support when considering the entire study. Whether this average group difference is significant is shown by the main effect for group in the mixed effects models.

Parameter estimates, standard errors and *p*-values for the time effect, overall HOPE effect (PAU serves as the reference category), and the HOPE by time interaction are shown in ***Exhibit 5-54*** for each outcome across sites and within each site.

Exhibit 5-53. Daily averages and best linear fit of attitudes by group, aggregated over sites

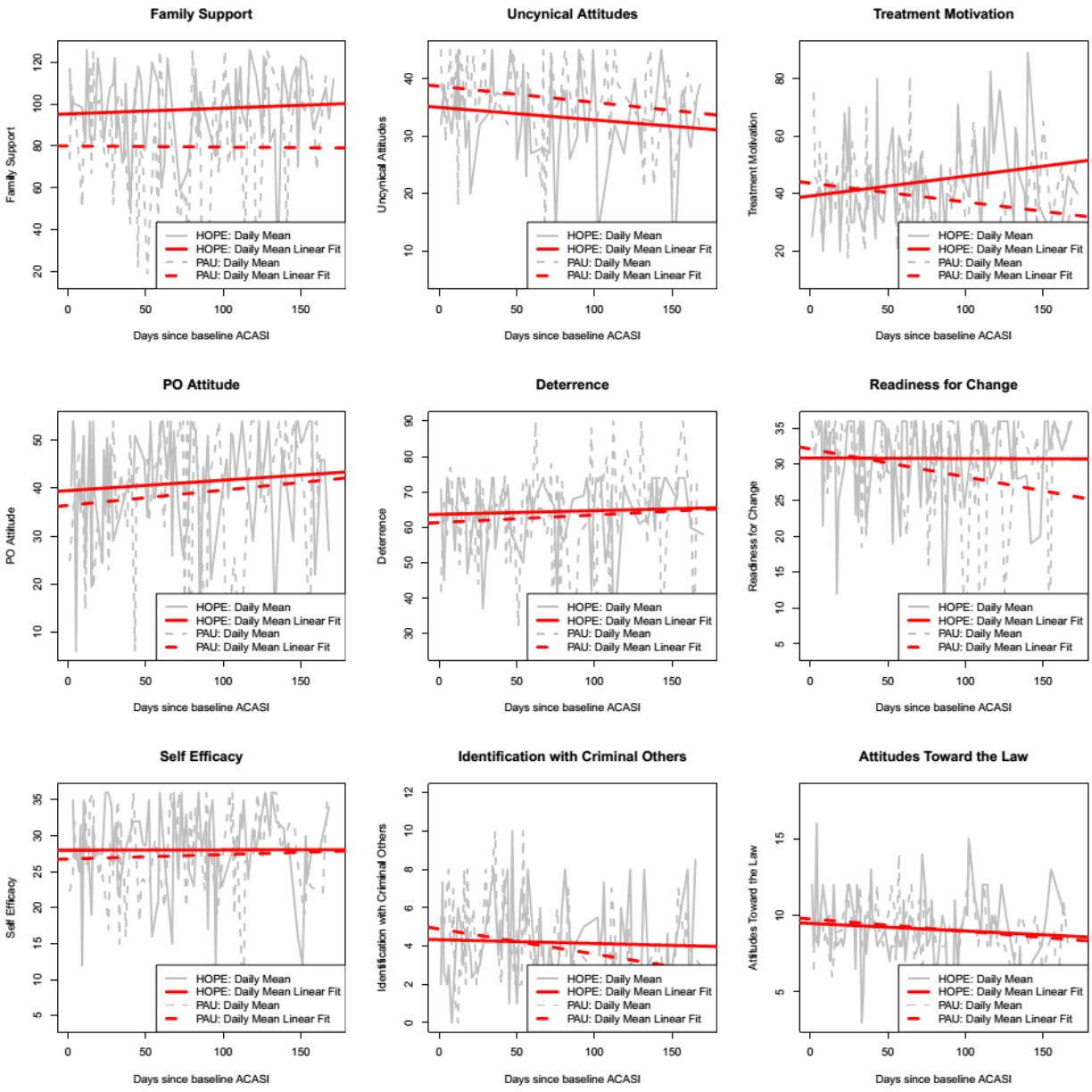


Exhibit 5-54. Parameter estimates from mixed effects repeated measures models of attitude change

Outcome	Site	Time			Group (PAU referent)			Group X Time		
		Estimate	SE	*p*	Estimate	SE	*p*	Estimate	SE	*p*
Family Support	**All[1]**	-0.03	0.04	0.45	**6.32**	**3.26**	**0.05**	*0.10*	*0.05*	*0.06*
	AR	-0.15	0.11	0.23	15.89	10.94	0.16	0.77	0.67	0.30
	MA	0.05	0.05	0.37	-6.20	5.04	0.22	-0.11	0.08	0.18
	OR	0.10	0.07	0.17	6.30	5.50	0.25	0.04	0.09	0.71
	TX	0.09	0.09	0.35	-0.99	7.67	0.90	-0.15	0.17	0.39
Uncynical Attitudes	**All[1]**	0.00	0.01	0.99	-0.59	1.00	0.56	-0.01	0.02	0.46
	AR	-0.02	0.03	0.59	0.40	2.94	0.89	0.19	0.14	0.20
	MA	0.00	0.02	0.86	-1.16	1.59	0.47	0.01	0.03	0.69
	OR	-0.02	0.02	0.36	-3.48	1.79	0.06	0.00	0.03	0.96
	TX	-0.04	0.03	0.21	0.66	2.22	0.77	*0.09*	*0.05*	*0.06*
Treatment Motivation	**All[2]**	**-0.06**	**0.03**	**0.03**	-1.81	1.95	0.35	**0.12**	**0.04**	**0.00**
	AR	**-0.12**	**0.05**	**0.05**	**-18.21**	**5.91**	**0.00**	0.08	0.20	0.71
	MA	**0.07**	**0.03**	**0.03**	-2.64	2.83	0.35	**-0.16**	**0.05**	**0.00**
	OR	-0.05	0.06	0.39	-4.66	3.86	0.23	0.07	0.08	0.40
	TX	0.05	0.05	0.30	-4.16	3.48	0.24	*0.16*	*0.07*	*0.06*
PO Attitude	**All[3]**	*0.04*	*0.02*	*0.09*	1.84	1.90	0.33	0.00	0.03	0.96
	AR	**0.13**	**0.06**	**0.05**	6.34	5.55	0.27	-0.18	0.10	0.11
	MA	0.05	0.03	0.11	-3.17	3.15	0.32	-0.03	0.05	0.52
	OR	0.02	0.04	0.70	-0.76	3.59	0.83	0.00	0.06	0.98
	TX	0.06	0.04	0.16	-1.18	3.34	0.72	0.09	0.07	0.23
Deterrence	**All[4]**	0.01	0.02	0.45	1.13	1.44	0.44	0.01	0.03	0.73
	AR	0.07	0.04	0.11	**-9.37**	**4.03**	**0.03**	0.76	0.45	0.14
	MA	0.02	0.03	0.53	-2.40	2.41	0.32	-0.07	0.04	0.08
	OR	0.03	0.03	0.39	1.18	2.09	0.57	-0.01	0.04	0.79
	TX	0.00	0.04	0.98	-0.93	2.91	0.75	**0.13**	**0.05**	**0.02**
Readiness for Change	**All[1]**	**-0.03**	**0.01**	**0.01**	-0.95	0.86	0.27	*0.03*	*0.01*	*0.08*
	AR	**-0.06**	**0.03**	**0.05**	*-4.42*	*2.42*	*0.08*	0.18	0.11	0.12
	MA	0.00	0.02	0.97	-0.04	1.48	0.98	-0.03	0.02	0.25
	OR	**-0.04**	**0.02**	**0.04**	*-2.53*	*1.40*	*0.07*	0.00	0.02	0.95
	TX	0.01	0.02	0.42	-2.03	1.52	0.19	0.01	0.03	0.80
Self-Efficacy	**All[5]**	0.00	0.01	0.84	-0.59	0.75	0.43	0.01	0.01	0.26
	AR	-0.03	0.03	0.34	2.65	2.68	0.33	0.34	0.27	0.23
	MA	*0.02*	*0.01*	*0.09*	0.20	1.19	0.87	-0.01	0.02	0.75
	OR	-0.02	0.01	0.28	**-2.92**	**1.19**	**0.02**	0.01	0.02	0.58
	TX	0.01	0.02	0.48	-1.03	1.63	0.53	-0.02	0.03	0.48

Outcome	Site	Time			Group (PAU referent)			Group X Time		
		Estimate	SE	*p*	Estimate	SE	*p*	Estimate	SE	*p*
Identification with Criminal Others	**All**[6]	*-0.01*	*0.00*	*0.10*	-0.33	0.27	0.22	*0.01*	*0.00*	*0.08*
	AR	-0.01	0.01	0.23	-0.93	0.75	0.22	-0.06	0.05	0.32
	MA	0.00	0.00	0.49	0.35	0.39	0.36	-0.01	0.01	0.44
	OR	0.00	0.01	0.51	-0.17	0.49	0.73	*0.02*	*0.01*	*0.06*
	TX	-0.01	0.01	0.20	0.04	0.74	0.96	-0.02	0.02	0.28
Attitudes Toward the Law	**All**[1]	**-0.01**	**0.00**	**0.00**	-0.12	0.24	0.62	0.00	0.00	0.70
	AR	**-0.03**	**0.01**	**0.00**	-0.15	0.64	0.81	-0.03	0.04	0.51
	MA	**-0.01**	**0.00**	**0.02**	0.10	0.37	0.79	0.00	0.01	0.66
	OR	-0.01	0.01	0.42	-0.11	0.45	0.80	0.00	0.01	0.62
	TX	**-0.02**	**0.01**	**0.00**	0.08	0.51	0.88	**0.03**	**0.01**	**0.02**

[1] The cross-site model controlled for site. The Type III tests of fixed effects for site were not significant.
[2] Site had a significant effect on treatment motivation, F(3,320)=3.39, p=.02. Planned contrasts show that Oregon had significantly higher treatment motivation than Arkansas, Massachusetts, and Texas.
[3] Site had a significant effect on PO attitudes, F(3,234)=3.76, p=.01. Planned contrasts show that Texas had significantly higher PO attitudes than Massachusetts and Oregon.
[4] Site had a significant effect on deterrence, F(3,294)=7.2, p=.003. Planned contrasts show that Massachusetts had significantly higher deterrence scores than Arkansas, while Texas had significantly higher deterrence scores than Massachusetts and Oregon.
[5] Site had a significant effect on self-efficacy, F(3,317)=4.86, p<.001. Planned contrasts show that Massachusetts, Oregon, and Texas had significantly higher self-efficacy than Arkansas.
[6] Site had a significant effect on identification with criminal others, F(3,306)=3.26, p=.02. Planned contrasts show that Oregon and Texas had significantly higher (i.e., lower scores) identification with criminal others than Massachusetts.

In the cross-site model for ***family support***, the overall time effect was not significant. On average, the HOPE group reported significantly more family support than the PAU group, and the HOPE group had a significant rate of improvement in family support (p=.06). These results are consistent with what is observed in ***Exhibit 5-50***. No effects were significant within site, suggesting a lack of power relative to the cross-site model.

In the cross-site model for ***Uncynical Attitudes,*** none of the effects is significant, consistent with the plot in ***Exhibit 5-53***, where the decreases over time and group differences are small. The group difference is significant in Oregon, where the effect is beta=-3.48, indicating lower (worse) attitudes on average in the HOPE group. This is consistent with the HOPE line being lower than the PAU line in ***Exhibit 5-53***. In Texas, there was improvement over time in these scores for the HOPE group.

In the model for ***Treatment Motivation, the overall time effect is negative, but the time effect for HOPE is positive suggesting improving attitudes towards treatment for the HOPE group relative to the PAU group.*** This is consistent with the lines going in opposite directions for treatment motivation in ***Exhibit 5-53***. ***In Arkansas, the overall trend is steeply deteriorating, and HOPE participants had significantly lower average treatment motivation (over time) than PAU participants. In Massachusetts, treatment motivation was improving overall, but at a slower rate for HOPE participants than PAU participants. Texas HOPE participants had improving treatment motivation***.

In the model for ***PO Attitudes***, the time effect is positive suggesting improving views of their probation officers' attitudes, with no group differences. This is consistent with the average trajectories in ***Exhibit 5-53***. This effect holds for Arkansas, and is in the same direction but is not significant for the other sites.

The ***Deterrence*** average trajectories were flat in ***Exhibit 5-53***, and the cross-site results confirm this observation. Deterrence scores deteriorated for HOPE participants in Massachusetts, but improved for HOPE participants in Texas.

On average, ***Readiness for Change*** deteriorated over time. Relative to the average change (-0.03), the HOPE group changed less negatively (0.03), yielding a null (-0.03 + 0.03 = 0) total change for HOPE, consistent with the average trajectories shown in ***Exhibit 5-53***. In Arkansas and Oregon, the HOPE group had lower scores than PAU, averaged over time.

In the model for ***Self-Efficacy***, the time effect is null, with no group differences. This is consistent with the self-efficacy average trajectories. In Massachusetts, self-efficacy increased overall, but without group differences.

Identification with Criminal Others decreased over time (suggesting less identification with criminal others), but at a slower rate for HOPE participants, consistent with the null (-.01 + 0.01 = 0) overall change seen in Exhibit 5-53 for HOPE. In Oregon, identification with criminal others became worse over time for HOPE participants.

On average, ***Attitudes toward the Law deteriorated over time, consistent with the downward slope in Exhibit 5-53***. The deterioration was larger in Arkansas and Texas, though HOPE participants in Texas had a net positive change in attitude toward the law (-0.02 + 0.03 = 0.01 increase per day).

The cross-site mixed models were also estimated using data from respondents who consented and called in at least once. Although p-values change, the direction and magnitudes of effects are generally consistent with those in ***Exhibit 5-54*** (results not shown). This is expected since there were no significant differences in attitudes when those who consented to T-ACASI but only completed the baseline were compared to those who called at least once, and graphically, cutting off day zero would do little to alter the graphical results. ***Importantly, for the evaluation, we saw relatively few significant changes over time in attitudes.***

5.7. Summary of Outcome Evaluation Findings

A total of 1,580 individuals were randomly assigned to HOPE (794) or to PAU (786) between August 2012 and September 2014. Of these, 76 individuals were determined after random assignment to be study ineligible, resulting in a final study sample of 743 HOPE probationers and 761 PAU probationers.[52] Most were male (81%), white (69%), and high risk (55%). On average, they were 31 years at study enrollment, with 7 prior arrests and 3.5 prior convictions. Most were on probation for either a drug (31%) or property offense (30%). Subject characteristics varied across the sites. For example, study participants were younger at first arrest in Texas than Arkansas (19 versus 27 years) and the average number of prior convictions ranged from 1.7 in Arkansas to 5.8 in Massachusetts.

Data for the outcome evaluation included administrative data collected from state and county agencies (e.g., arrests, revocations, convictions, probation violations, drug test results); drug test results

[52] 68 individuals were program ineligible; 8 were randomized twice and were retained in their original study and program assignment.

collected as part of the HOPE fidelity assessment process; and interview data collected at study intake, and at 6 and 12 months following intake.[53]

Recidivism outcomes were similar for the HOPE and PAU groups: 40% of HOPE versus 44% of PAU had a new arrest; 25% of HOPE versus 22% of PAU had a revocation; 49% of HOPE versus 50% of PAU had an arrest or revocation; and 28% of HOPE versus 26% of PAU had a new conviction. There was some variation in rates across sites, but the general conclusions of no differences hold with a few exceptions:

1. HOPE probationers were more likely than PAU probationers to be revoked in Arkansas and Oregon, although PAU revocation rates in those sites were low—about 10%--suggesting little opportunity for lower rates for those on HOPE probation; and
2. HOPE probationers were more likely to have a new conviction in Arkansas.
3. HOPE probationers overall were less likely to have a new property charge than PAU (15% versus 20%) and new drug charge (12% versus 15%).
4. HOPE probationers in Texas were significantly less likely to have experienced a recidivism drug charge.

Lognormal survival models of time to recidivism events (arrest, arrest or revocation, new conviction) confirm the bivariate findings, but revealed one additional finding. Although time to revocation was less for HOPE probationers in Arkansas and Oregon, consistent with the higher observed rates, HOPE probationers had longer times to revocation in Texas.

Parameter estimates from models that examined predictors of the number (count) of new arrests showed no significant differences for HOPE versus PAU[54]. In the overall model, the coefficient for the site indicator for Arkansas was significant and positive—suggesting more arrests in Arkansas compared to Texas (reference category). Males were also predicted to have more arrests, overall and in the Texas model. Results were similar when the outcome was number of new arrest charges—although in the overall model, Oregon also had significantly higher counts than Texas.

HOPE was to hold individuals accountable to their supervision conditions, including compliance with intensive random drug testing—suggesting that ***HOPE probationers would have more violations which is what was observed.*** HOPE probationers were more likely to have a violation (89% versus 82%) and had more violations than PAU probationers (3,770 versus 3,134)—mostly drug-related violations (2,107 versus 915) attributable to testing (26,991 for HOPE versus 4,942 tests for PAU probationers). ***HOPE probationers were less likely to miss a probation officer visit (30% versus 44%), to fail to pay their fees and fines (11% versus 18%), and to be violated for a new charge (22% versus 28%). HOPE probationers were more likely to have a violation for failing to appear for court (18% versus 6%), although this may be because they had more hearings***. Most sanctions for HOPE probationers were jail (2,920 of 3,550 sanctions) and HOPE probationers were more likely to go to jail (82% versus 56%), to go more often (3.8 stays versus 1.4), and to serve more days total (47 versus 33.3 days) with a median stay of 4 days.

HOPE included treatment referral after repeated failed tests; HOPE participants were more likely to go to residential treatment (33% versus 11%). HOPE probationers were also referred to treatment more

[53] Response bias analyses suggested no differences between those who were interviewed and those who were not interviewed at any wave(Section 2.3).

[54] Massachusetts was excluded from these analyses because of data limitations.

quickly (overall and in three sites). ***Drug tests conducted in conjunction with follow-up interviews showed fewer positives for HOPE than PAU probationers***.

Interview findings showed the following:

- Probationers on HOPE and PAU reported similar rates of employment and similar wages across the three waves of interviews.
- HOPE probationers were more likely to have a job with formal pay at the 12-month interview.
- Both groups reported emotional problems. Overall HOPE probationers reported a lower average mental health symptom level at 12 months. There were no differences at any interview wave in self-assessment of emotional problems interfering with work or other activities. About 40% of both groups reported at all waves that they had accomplished less than they would have liked because of emotional problems.
- Both groups agreed that they needed education and were neutral with respect to needing job training.
- Neither group believed that they needed substance abuse or mental health treatment.
- There were no differences between the groups on the receipt of education and employment services.
- HOPE probationers were more likely than PAU probationers to report having received residential substance abuse treatment at the 6- and 12-month interviews.
- At the 12-month interview, HOPE probationers were much less likely than PAU probationers to report that most or all close friends are frequently drunk or high (10% versus 20%) or that most or all close friends have been incarcerated (20% versus 31%).

Thus, outcome results are consistent with much of the underlying HOPE theory of change: HOPE participants were extensively subjected to random drug testing and received many more violations as a result of failing tests or failing to appear for tests. The HOPE participants also experienced many more (short) jail stays—again as would have been expected given the underlying model. HOPE participants were also much more likely to be sent to residential drug treatment and to have a shorter time to referral. The interview data suggested that there was some attitudinal changes over time, such as in self-efficacy, that may support improvements in self-management as was discussed previously in the Process Evaluation section. Contrary to the original Hawaii HOPE findings, however, the overall conclusion is that HOPE did not impact criminal recidivism.

6. Findings from the Cost Study

The cost study focused on identifying the costs of starting and implementing HOPE programs and identifying the costs and any savings attributable to HOPE programs compared to PAU. Our general approach was to examine per-person costs. The methods for the cost study were described in section 2.4. Quantities and prices for intake, warning hearing, staff meetings, office visits, drug tests, violation hearings, arrests, state and county corrections, state prison, and residential treatment were included in the analyses. Costs were assessed for three periods, based on available data for subsets of our study participants determined by their (minimum) time in the study—monthly costs over a 24-month period, monthly costs over a 12-month period, and monthly costs over a 6-month period. The 24-month analyses include those in the study 24 months or longer. The 12-month and 6-month analyses include larger numbers of study subjects who were in the study 12 months or longer and 6 months or longer.

Unit price estimates are shown in ***Exhibit 6-1***. (Sources for the price data are shown in ***Exhibit 2-24;*** information about procedures is provided in section 2.4.) Highest unit costs are for arrests; but, prison, jail, and residential costs, which are per-day costs, can also add up. There is also considerable variation in the daily costs of prison and jails, with prison costs ranging from $57/day in Texas to $220/day in Massachusetts and jail costs ranging from $30/day in Arkansas to $94/day in Oregon. Residential treatment daily costs and arrest costs are from the literature and are constant across the sites.

Interviews were used to establish costs for the HOPE-specific intervention events. There was some variation across the sites, mostly attributable to different estimates of the staff time and, in some cases, the frequency of each event. For example, the estimated cost of a warning hearing ranged from $18 in Oregon to $51 in Texas and the estimated cost of a violation hearing ranged from $24 in Texas to $56 in Massachusetts.

Exhibit 6-1. Unit prices for the cost analyses

	AR		MA		OR		TX	
	PAU	HOPE	PAU	HOPE	PAU	HOPE	PAU	HOPE
Prison	$64.42	$64.42	$219.71	$219.71	$92.31	$92.31	$56.91	$56.91
State & county corrections	$29.72	$29.72	$91.21	$91.21	$93.64	$93.64	$65.47	$65.47
Residential treatment	$79.99	$79.99	$79.99	$79.99	$79.99	$79.99	$79.99	$79.99
Arrests	$1,385.00	$1,385.00	$1,385.00	$1,385.00	$1,385.00	$1,385.00	$1,385.00	$1,385.00
Intake	$29.33	$23.74	$48.80	$48.80	$34.36	$22.55	$36.07	$36.07
Office visits	$15.36	$13.96	$31.19	$27.12	$22.75	$16.85	$10.42	$14.02
Drug tests	$4.67	$8.16	$6.40	$6.40	$4.00	$4.00	$27.80	$27.80
Warning hearings	$0.00	$28.87	$0.00	$35.63	$0.00	$18.38	$0.00	$51.26
Violation hearings	$31.67	$31.67	$56.28	$56.28	$34.13	$34.13	$24.35	$24.35
Staffing meetings	$0.00	$3.28	$0.00	$0.00	$0.00	$28.62	$0.00	$3.70

Note: Sources are described in section 2.4 and Exhibit 2-24.

6.1. 24-Month Cost Analysis

Exhibit 6-2 shows the average cumulative monthly costs per probationer for HOPE and PAU, for probationers in the study at least 24 months. Although the costs and cost trajectories vary considerably by site and duration, the per-probationer cost of HOPE is always greater than PAU. Cost differences between the two study arms also tend to increase over time.

Exhibit 6-2. Average cumulative costs per probationer by site and month.

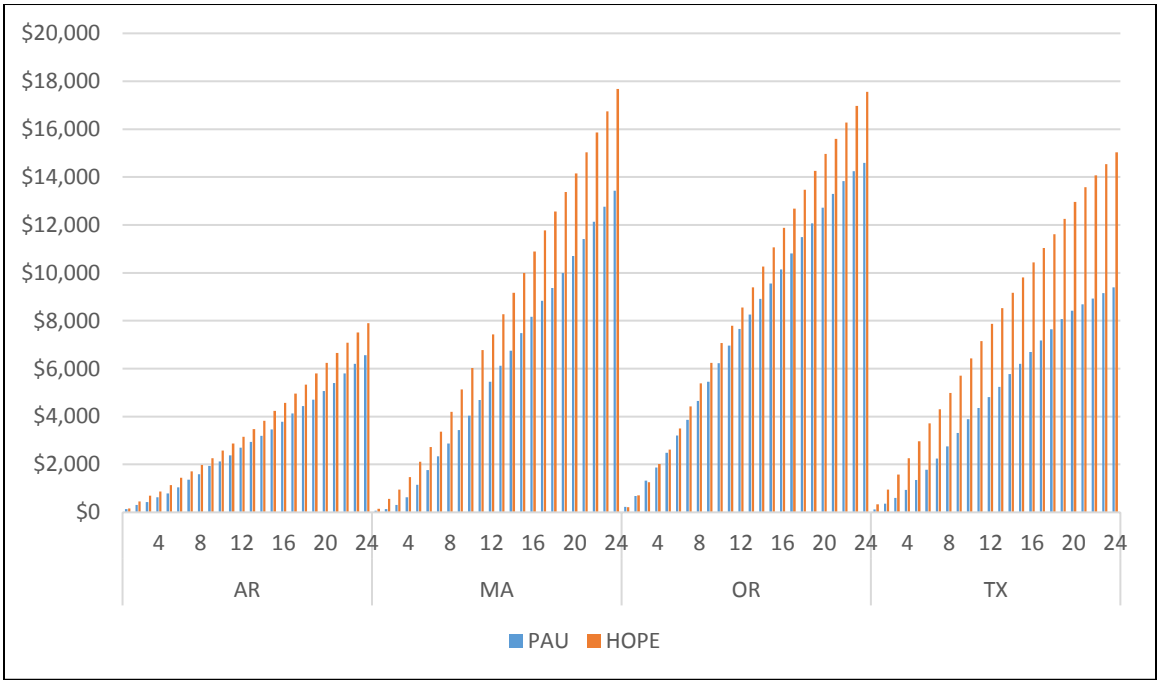


Exhibit 6-3 shows mean and median total costs for HOPE and PAU for each of the sites and aggregated across sites. For each site and overall, ***the mean and median HOPE cost is always higher than the mean and median cost for PAU.*** The table also shows t-test results comparing mean total costs between HOPE and PAU. The difference is statistically significant at the 0.05 level for one of the four sites individually, and when comparing average costs across all four sites in aggregate the difference is significant.

In all instances, the mean is substantially higher than the median, suggesting there are some high-cost probationer outliers. Non-normally distributed data can bias t-test estimates so we also conducted a non-parametric equality-of-medians test (two-sample equality of medians test, distributed chi-square with 1 degree of freedom) that compares the difference in medians across treatment groups. The specific results of the median test vary slightly from the t-tests, but the overall finding does not change: differences between HOPE and PAU are significant for only one site individually, but for all sites in aggregate.

Exhibit 6-3. Average total costs for 24 months of supervision per probationer by site.

Site	PAU			HOPE			t test	Median Test
	N	Mean	Median	N	Mean	Median		
AR	54	$6,563	$2,139	68	$7,901	$3,184	0.7	0.1
MA	61	$13,425	$2,721	56	$17,672	$6,727	1.0	3.8
OR	103	$14,588	$9,600	96	$17,564	$14,015	1.4	3.1
TX	93	**$9,392**	**$5,261**	94	**$15,038**	**$13,799**	**4.1*** **	**4.5***
All	*311*	***$11,413***	***$5,797***	*314*	***$14,735***	***$10,355***	***2.7****	***7.6****

*p < 0.01**p < 0.01; *** p < 0.001

Exhibit 6-4 shows the average total cost by category for each group by site for the 24-month sample. Average prison costs are higher for PAU probationers in two sites (Massachusetts and Texas) and lower in two sites (Arkansas and Oregon). Jail costs are similar for the two groups in Arkansas and lower in the other three sites. Residential treatment costs are higher for the HOPE groups in all four sites. Arrest costs are higher for PAU in Arkansas and Texas and lower in Massachusetts. Intake costs were similar for both groups.

Exhibit 6-4. Average cost by category, site, and group (24-month sample)

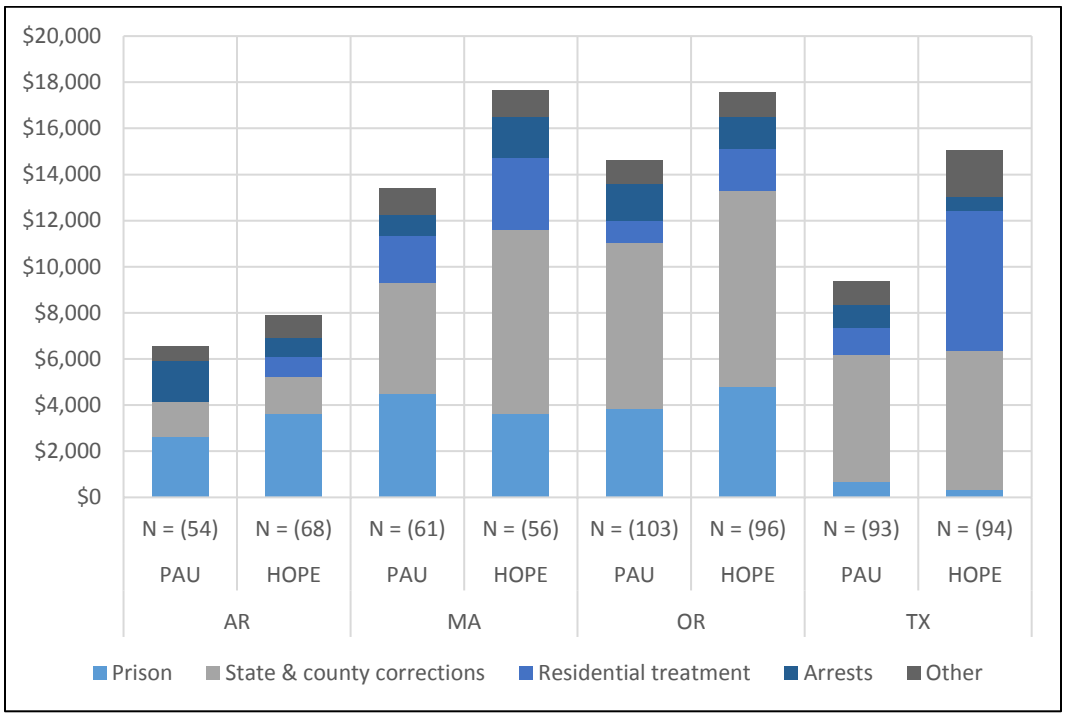
	AR		MA		OR		TX	
	PAU	HOPE	PAU	HOPE	PAU	HOPE	PAU	HOPE
N	54	68	61	56	103	96	93	94
Prison	$2,620.94	$3,638.78	$4,455.43	$3,633.06	$3,844.76	$4,782.81	$692.10	$347.51
Jail[1]	$1,523.43	$1,592.21	$4,858.05	$8,005.31	$7,204.83	$8,532.94	$5,488.92	$6,026.72
Res. Tx.	$0.00	$879.89	$2,010.24	$3,082.47	$941.24	$1,784.78	$1,179.21	$6,046.05
Arrests	$1,795.37	$835.07	$953.61	$1,780.71	$1,600.15	$1,385.00	$1,012.69	$589.36
Intake	$29.33	$23.74	$48.80	$48.80	$34.36	$22.55	$36.07	$36.07
Office visits	$440.44	$382.26	$943.65	$763.32	$729.13	$506.74	$517.97	$623.50
Drug tests	$41.25	$409.56	$18.57	$94.51	$26.41	$154.79	$317.76	$1,117.62
Warning hearings	$0.00	$24.15	$0.00	$35.72	$0.00	$18.20	$0.00	$40.08
Violation hearings	$112.60	$101.53	$136.55	$228.13	$207.10	$204.42	$147.41	$183.66
Staffing meetings	$0.00	$13.99	$0.00	$0.00	$0.00	$171.42	$0.00	$27.91
Total	*$6,563.36*	*$7,901.18*	*$13,424.90*	*$17,672.05*	*$14,587.96*	*$17,563.66*	*$9,392.12*	*$15,038.49*

[1] State and local corrections

For events that were highlighted by drug supervision, we see that average drug testing costs were higher for HOPE in all sites—driven primarily by the substantial differences in numbers of tests of the HOPE and PAU groups. The average total costs for violation hearings were similar for HOPE and PAU in Arkansas and Oregon, and somewhat higher for the HOPE groups in Massachusetts and Texas.

The average total costs for HOPE and PAU by costs attributed to prison, state and county corrections, residential treatment, arrest, and other (including intake, office visits, drug tests, warning hearings, violation hearings, and staffing meetings) are shown in ***Exhibit 6-5***. ***Incarceration and residential treatment drive total costs for both HOPE and PAU groups and the cost differential between these groups.*** Arrest and other cost categories contribute minimally to total costs and vary relatively little by treatment arm and site.

Exhibit 6-5. Distribution of average costs across events by site



6.2. 12-Month and 6-Month Cost Sub-Analyses

Results for more participants observed for 12- and 6-month periods are presented in this section. At the expense of a shorter period, these sub-analyses include more observations that allow for better hypothesis testing. In addition, comparing the 6, 12, and 24-month samples allows us to identify patterns over time.

Exhibits 6-6 and ***6-7*** show the average cumulative monthly costs per probationer for HOPE and PAU for the 6- and 12-month samples. These figures show that the main result is borne out after altering the period and adding more observations: HOPE is costlier than PAU across all sites and in every month. As with the primary analyses, the cost difference generally increases over time.

Exhibit 6-6. Average cumulative costs by site and month, 6-month sample.

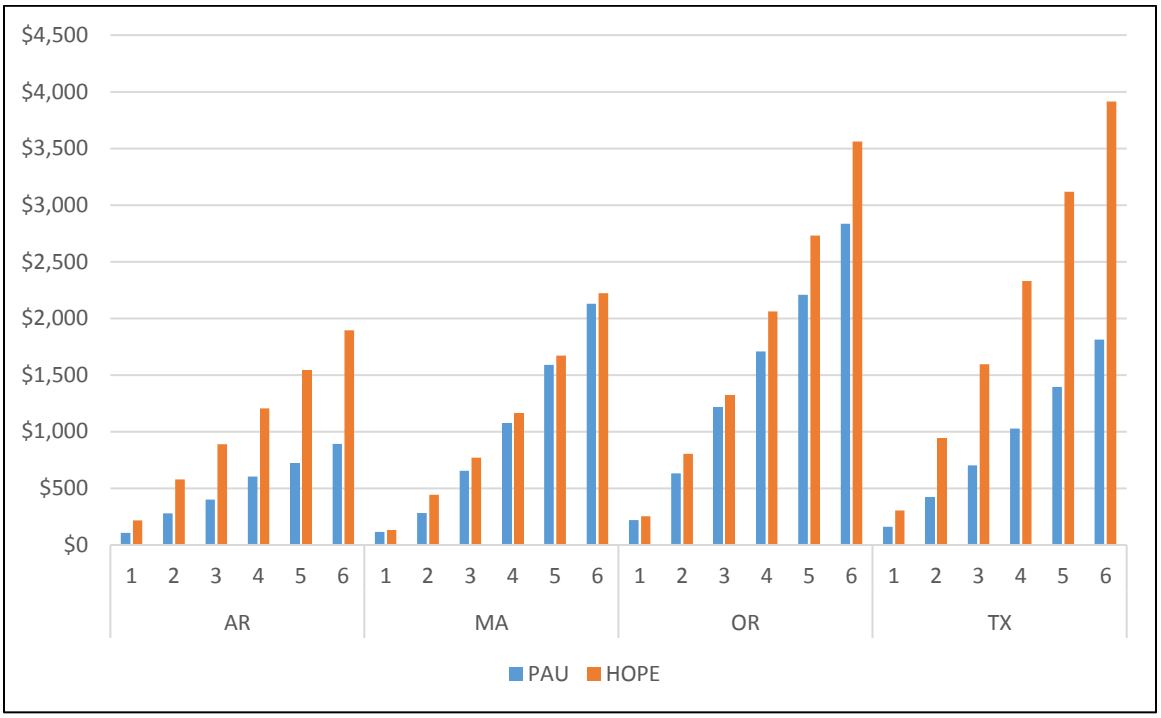
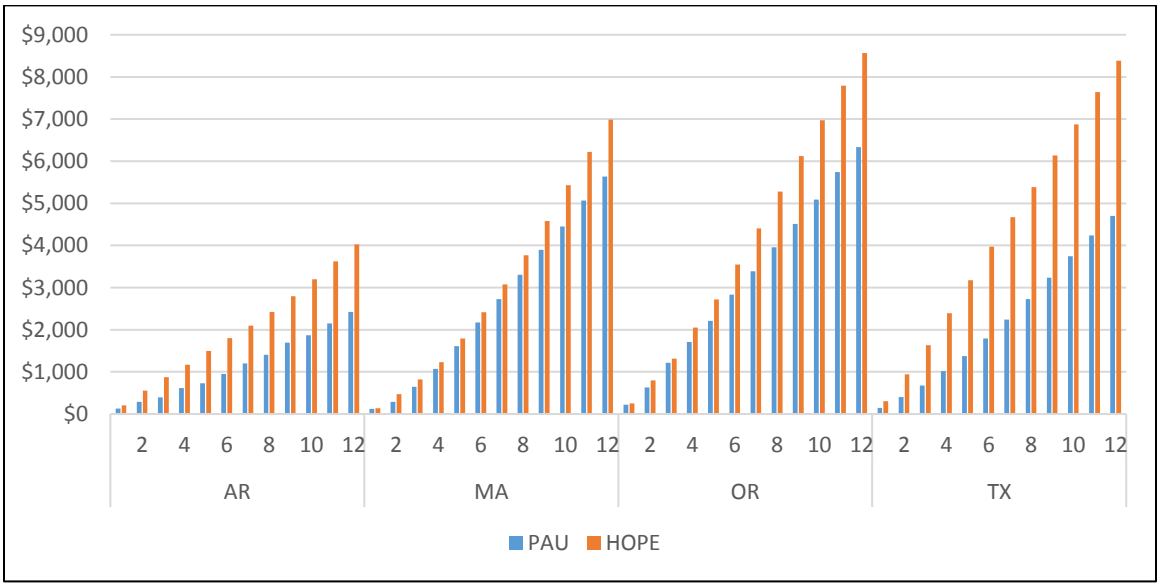


Exhibit 6-7. Average cumulative costs by site and month, 12-month sample.



Exhibits 6-8 and *6-9* show mean and median total costs for HOPE and PAU for each of the sites and aggregated across sites, for the 6- and 12-month samples. Again, we see that in nearly all instances, mean and median HOPE costs are higher than PAU. However, with these larger sample sizes we also see that in more instances the difference is statistically significant.

Exhibit 6-8. Average total costs per probationer by site, 6-month sample

Site	PAU			HOPE			t test	Median test
	N	Mean	Median	N	Mean	Median		
AR	**159**	**$892**	**$190**	**179**	**$1,893**	**$1,038**	**5.4*****	**56.5*****
MA	199	$2,128	**$329**	188	$2,223	**$651**	0.3	**30.7*****
OR	203	**$2,836**	$1,162	190	**$3,562**	$1,983	**2.0***	3.5
TX	191	**$1,813**	**$639**	185	**$3,913**	**$2,723**	**7.3*****	**47.8*****
All	*752*	***$1,978***	***$364***	*742*	***$2,908***	***$1,698***	***5.8*****	***84.8*****

*p < 0.05; **p < 0.01; *** p < 0.001

Exhibit 6-9. Average total costs per probationer by site, 12-month sample.

Site	PAU			HOPE			t test	Median test
	N	Mean	Median	N	Mean	Median		
AR	112	**$2,420**	**$418**	134	**$4,028**	**$1,939**	**3.0*****	**4.2***
MA	167	$5,637	**$733**	153	$6,984	**$2,104**	1.3	**6.6***
OR	203	$6,333	$3,826	189	$8,566	$5,044	**2.9*****	3.0
TX	167	**$4,700**	**$2,150**	166	**$8,386**	**$8,718**	**6.1*****	**26.0*****
All	*649*	***$5,059***	***$1,863***	*642*	***$7,195***	***$4,015***	***5.4*****	***36.5*****

*p < 0.05; **p < 0.01; *** p < 0.001

Exhibits 6-10 and *6-11* decompose the average total costs for HOPE and PAU by the costs attributed to each criminal justice event for the 6- and 12-month samples. As with the primary analyses, ***incarceration and residential treatment drive total costs for both treatment arms and are the primary reason for increased cost under HOPE.*** However, looking across the three sample periods these sub-analyses also indicate that as time increases, incarceration and residential treatment contribute more to total costs while the arrests and other events contribute less. The average total costs for HOPE and PAU by costs attributed to prison, state and county corrections, residential treatment, arrest, and other (including intake, office visits, drug tests, warning hearings, violation hearings, and staffing meetings) are shown in ***Exhibit 6-12*** and ***6-13***.

Exhibit 6-10. Average cost by category, site, and group (6-month sample)

	AR		MA		OR		TX	
	PAU	HOPE	PAU	HOPE	PAU	HOPE	PAU	HOPE
N	159	179	198	188	203	190	191	185
Prison	$42.54	$206.22	$199.84	$7.01	$264.20	$473.70	$20.86	$13.54
Jail[1]	$348.04	$576.63	$893.31	$1,223.09	$1,525.46	$2,082.75	$705.09	$1,303.38
Res. Tx.	$0.00	$387.44	$378.65	$280.82	$286.47	$251.34	$452.30	$1,588.56
Arrests	$331.01	$348.18	$327.11	$287.31	$470.76	$379.05	$311.81	$232.08
Intake	$29.33	$23.74	$48.80	$48.80	$34.36	$22.55	$36.07	$36.07
Office visits	$109.94	$90.93	$233.65	$200.46	$180.82	$125.52	$128.76	$146.49
Drug tests	$6.87	$161.29	$7.91	$52.83	$7.53	$64.25	$91.55	$447.66
Warning hearings	$0.00	$27.86	$0.00	$35.63	$0.00	$18.27	$0.00	$48.93
Violation hearings	$24.70	$63.52	$39.03	$87.11	$66.58	$78.68	$66.55	$83.71
Staffing meetings	$0.00	$7.66	$0.00	$0.00	$0.00	$65.98	$0.00	$12.72
Total	*$892.43*	*$1,893.46*	*$2,128.29*	*$2,223.07*	*$2,836.17*	*$3,562.08*	*$1,812.97*	*$3,913.13*

[1] State and local corrections

Exhibit 6-11. Average cost by category, site, and group (12-month sample)

	AR		MA		OR		TX	
	PAU	HOPE	PAU	HOPE	PAU	HOPE	PAU	HOPE
N	112	134	166	153	203	189	168	166
Prison	$272.06	$1,465.80	$801.22	$967.87	$899.91	$1,788.57	$95.76	$139.88
Jail[1]	$795.28	$854.12	$2,840.07	$3,620.38	$3,516.34	$4,459.54	$2,622.33	$2,676.38
Res. Tx.	$0.00	$627.98	$791.76	$1,076.47	$497.67	$1,034.37	$810.44	$3,943.60
Arrests	$1,026.38	$465.11	$605.42	$633.66	$873.30	$652.20	$580.54	$367.11
Intake	$29.33	$23.74	$48.80	$48.80	$34.36	$22.55	$36.07	$36.07
Office visits	$226.31	$187.75	$469.05	$390.50	$370.41	$248.69	$254.84	$296.28
Drug tests	$16.22	$280.00	$12.30	$74.00	$14.25	$102.58	$192.60	$728.49
Warning hearings	$0.00	$27.26	$0.00	$34.98	$0.00	$18.55	$0.00	$47.11
Violation hearings	$54.86	$86.27	$68.75	$137.57	$126.43	$129.84	$107.61	$131.43
Staffing meetings	$0.00	$10.42	$0.00	$0.00	$0.00	$108.88	$0.00	$19.97
Total	*$2,420.43*	*$4,028.44*	*$5,637.37*	*$6,984.24*	*$6,332.68*	*$8,565.76*	*$4,700.18*	*$8,386.32*

[1] State and local corrections

Exhibit 6-12. Distribution of costs across events by site and group, 6-month sample

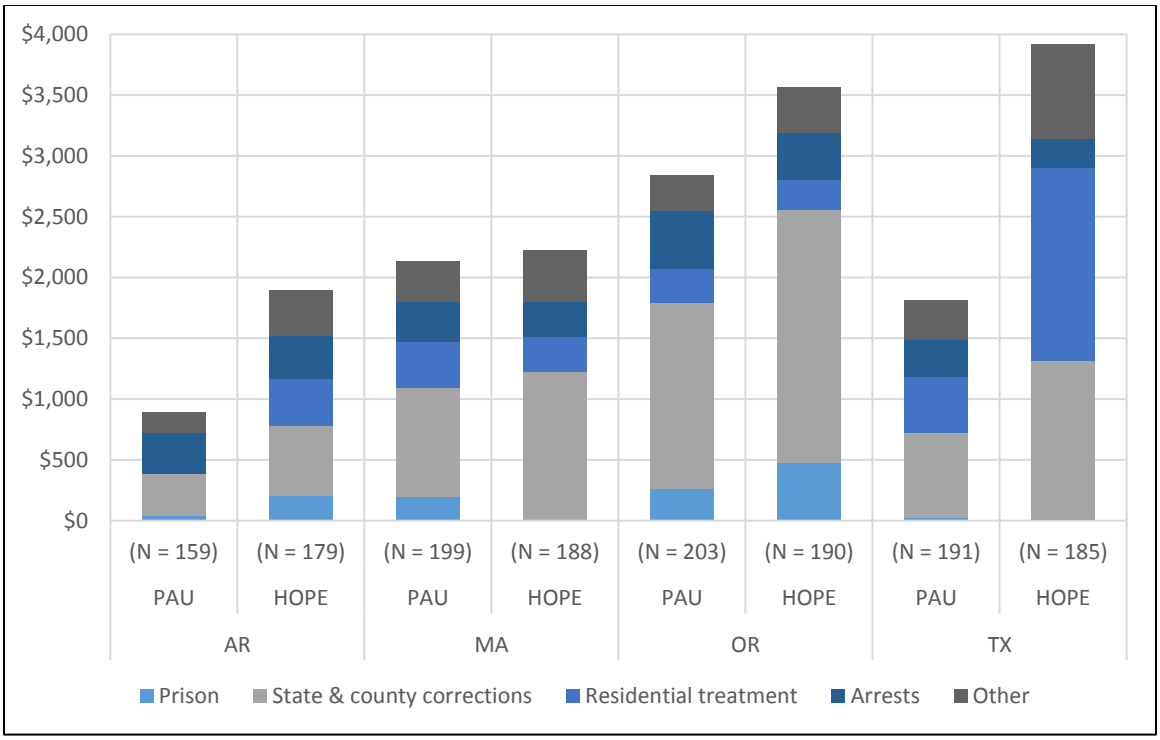
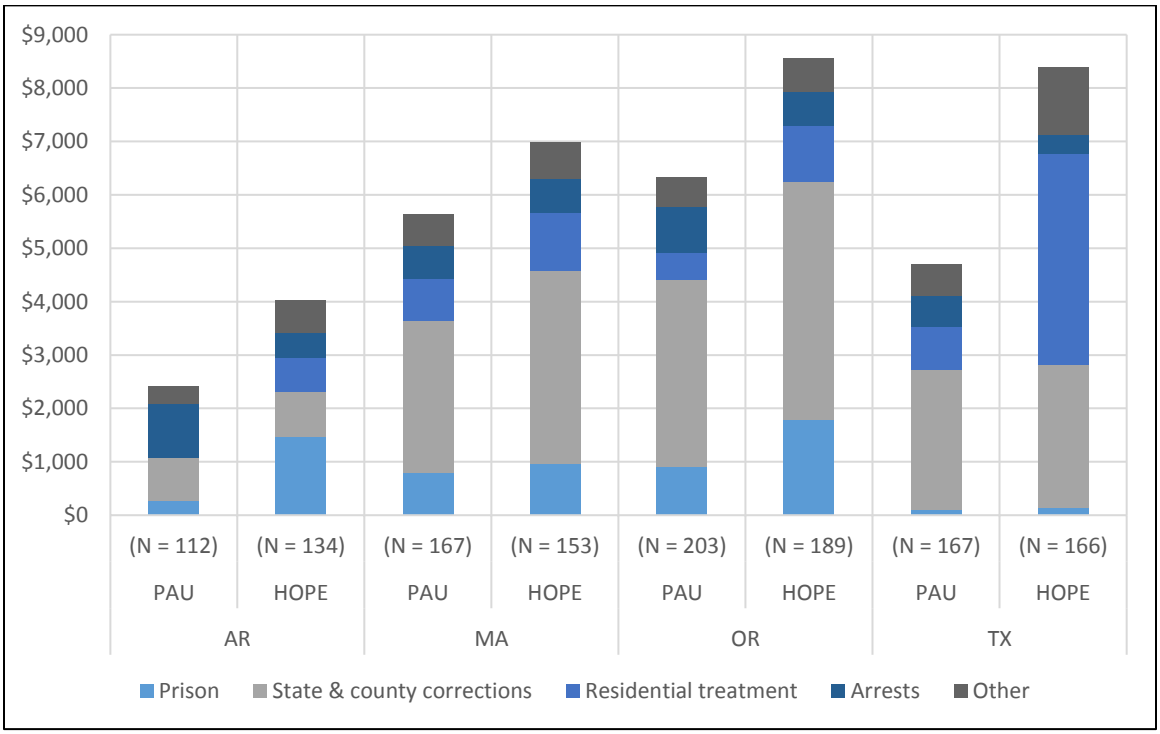


Exhibit 6-13. Distribution of costs across events by site and group, 12-month sample



6.3. Cost Sensitivity Analyses

We conducted three sensitivity analyses to explore the effect of study assumptions on the conclusions. The first two sensitivity analyses consider how a change in the assigned incarceration price effects the study results. In Sensitivity Analysis 1, we assigned to all sites the value of the cost per night in prison that is lowest across sites, as well as the lowest cost per night in jail. In Sensitivity Analysis 2, we assigned to all the sites the cost per night in prison that is highest across sites, as well as the highest cost per night in jail. Sensitivity Analysis 3 considers how a change in the assigned arrest price effects the study results, and we do this by assigning $0.00 as the arrest price across all sites. Results are shown in ***Exhibits 6-14***, ***6-15***, and ***6-16***.

The broad study conclusions that HOPE costs more, that cost differences rise over time, and that incarceration and treatment drive costs did not change under any of these sensitivity analyses. All sites and sample sizes that were significant at the 0.05-level in the primary analyses remained significant under all three sensitivity analyses, with five exceptions. Specifically, under Sensitivity Analysis 1, the comparison for the mean costs between HOPE and PAU for the Oregon 6-month sample "lost" statistical significance as the p-value of the t-statistic decreased from 0.04 to 0.23.

We saw three changes in significance under Sensitivity Analyses 2. The differences in mean costs between HOPE and PAU were no longer significant for the Oregon 6-month sample (p-values decreased from 0.04 to 0.06) and for the Texas 24-month sample (p-values decreased from 0.00 to 0.06) and the differences in median costs were no longer significant for the Texas 24-month sample (p-values decreased from 0.03 to 0.07).

Under Sensitivity Analyses 3, we saw one change. The difference in median costs between HOPE and PAU attained significance as the p-value increased from 0.06 under the primary analyses to 0.01 under the Sensitivity Analysis.

The primary analysis results show—and sensitivity analysis results confirm— that cost differences by treatment group are driven by quantity, not price. Per unit prices assigned to the largest categories of costs (prison, jail, residential treatment, and arrests) do not vary between HOPE and PAU. Therefore, increased costs across treatment group for these categories are driven by differences in the number of events, especially the number of nights in prison, jail, and residential treatment, and the average number of arrests. Because HOPE incurs higher average quantities of these events, it incurs higher costs.

Exhibit 6-14. Sensitivity analysis 1: Low cost per night incarcerated

	PAU			HOPE			Means Test		Medians Test	
	N	Mean	Median	N	Mean	Median	t	p value	Chi sq	p value
6-month sample										
AR	159	**$887**	**$190**	179	**$1,869**	**$1,038**	**-5.4**	**0.00**	**56.5**	**0.00**
MA	199	$1,378	$**329**	188	$1,393	**$551**	-0.1	0.94	**30.7**	**0.00**
OR	203	$1,694	$615	190	$1,959	$992	-1.2	0.23	3.5	0.06
TX	191	**$1,428**	**$492**	185	**$3,201**	**$1,941**	**-7.0**	**0.00**	**63.1**	**0.00**
All	*752*	***$1,372***	***$357***	*742*	***$2,104***	***$1,147***	***-6.5***	***0.00***	***109.3***	***0.00***
12-month sample										
AR	112	**$2,389**	**$418**	134	**$3,858**	**$1,939**	**-2.9**	**0.00**	**4.2**	**0.04**
MA	167	$3,129	**$717**	153	$3,826	**$1,265**	-1.4	0.16	**5.5**	**0.02**
OR	203	**$3,587**	$2,186	189	**$4,836**	$2,727	-2.7	0.01	0.8	0.36
TX	167	**$3,268**	**$1,642**	166	**$6,925**	**$6,976**	**-7.6**	**0.00**	**23.8**	**0.00**
All	*649*	***$3,180***	***$1,413***	*642*	***$4,931***	***$2,809***	***-7.0***	***0.00***	***26.5***	***0.00***
24-month sample										
AR	54	$6,258	$2,139	68	$7,477	$3,184	-0.7	0.46	0.1	0.72
MA	61	$6,848	$2,642	56	$9,583	$4,674	-1.5	0.14	2.5	0.12
OR	103	$8,195	$4,614	96	$9,905	$6,555	-1.3	0.19	2.2	0.14
TX	93	$6,395	$4,617	94	$11,748	$10,773	**-5.0**	**0.00**	**19.9**	**0.00**
All	*311*	*$7,056*	*$3,695*	*314*	*$9,873*	*$6,555*	***-4.0***	***0.00***	***7.6***	***0.01***

Exhibit 6-15. Sensitivity analysis 2: High cost per night incarcerated

	PAU			HOPE			Means Test		Medians Test	
	N	Mean	Median	N	Mean	Median	t	p value	Chi sq	p value
6-month sample										
AR	159	**$1,744**	**$197**	179	**$3,631**	**$1,879**	**-4.3**	**0.00**	**44.2**	**0.00**
MA	199	$2,152	**$329**	188	$2,256	**$654**	-0.3	0.77	**30.7**	**0.00**
OR	203	$3,201	$1,162	190	$4,216	$1,983	-1.9	0.1	3.5	0.06
TX	191	**$2,176**	**$777**	185	**$4,513**	**$3,456**	**-6.9**	**0.00**	**53.6**	**0.00**
All	*752*	***$2,355***	***$366***	*742*	***$3,652***	***$1,993***	***-6.0***	***0.00***	***79.2***	***0.00***
12-month sample										
AR	112	**$4,787**	**$426**	134	**$9,399**	**$3,434**	**-2.9**	**0.00**	**11.1**	**0.00**
MA	167	$5,713	**$733**	153	$7,081	**$2,108**	-1.3	0.19	**6.6**	**0.01**
OR	203	$7,575	$4,033	189	$11,034	$5,044	-2.6	0.01	3.0	0.09
TX	167	**$6,102**	**$2,627**	166	**$9,938**	**$9,712**	**-4.4**	**0.00**	**23.8**	**0.00**
All	*649*	***$6,236***	***$2,030***	*642*	***$9,467***	***$4,783***	***-5.3***	***0.00***	***32.6***	***0.00***
24-month sample										
AR	54	$16,158	$3,473	68	$20,097	$4,725	-0.7	0.49	0.1	0.72
MA	61	$13,554	$2,721	56	$17,885	$6,830	-1.0	0.32	3.8	0.05
OR	103	$19,894	$9,787	96	$24,165	$14,675	-1.0	0.33	3.1	0.08
TX	93	$13,734	$6,354	94	$18,626	$15,433	-1.9	0.1	3.3	0.07
All	*311*	***$16,160***	***$6,240***	*314*	***$20,506***	***$11,647***	***-2.1***	***0.04***	***8.5***	***0.00***

Exhibit 6-16. Sensitivity analysis 3: No cost per arrest

	PAU			HOPE			Means Test		Medians Test	
	N	Mean	Median	N	Mean	Median	t	p value	Chi sq	p value
6-month sample										
AR	159	**$561**	**$183**	179	**$1,545**	**$784**	**-6.6**	**0.00**	**98.3**	**0.00**
MA	199	$1,801	**$328**	188	$1,936	**$625**	-0.4	0.67	**40.4**	**0.00**
OR	203	**$2,365**	**$959**	190	**$3,183**	**$1,869**	**-2.5**	**0.01**	**7.1**	**0.0**
TX	191	**$1,501**	**$573**	185	**$3,681**	**$2,390**	**-8.0**	**0.00**	**84.3**	**0.00**
All	*752*	***$1,615***	***$336***	*742*	***$2,596***	***$1,350***	***-6.7***	***0.00***	***139.2***	***0.00***
12-month sample										
AR	112	**$1,394**	**$378**	**134**	**$3,563**	**$1,550**	**-4.6**	**0.00**	**34.7**	**0.00**
MA	167	$5,032	**$693**	153	$6,351	**$1,703**	-1.4	0.18	**19.0**	**0.00**
OR	203	**$5,459**	**$3,231**	189	**$7,914**	**$4,810**	**-3.5**	**0.00**	**3.7**	**0.05**
TX	167	**$4,120**	**$1,615**	166	**$8,019**	**$8,624**	**-6.8**	**0.00**	**33.1**	**0.00**
All	*649*	***$4,303***	***$1,127***	*642*	***$6,660***	***$3,502***	***-6.3***	***0.00***	***57.7***	***0.00***
24-month sample										
AR	54	$4,768	**$834**	68	$7,066	**$2,146**	-1.4	0.18	**3.3**	**0.07**
MA	61	$12,471	$1,489	56	$15,891	$4,575	-0.8	0.42	1.4	0.23
OR	103	$12,988	$8,215	96	$16,179	$11,547	-1.5	0.13	2.2	0.14
TX	93	**$8,379**	**$4,040**	94	**$14,449**	**$13,111**	**-4.6**	**0.00**	**4.5**	**0.03**
All	*311*	***$10,081***	***$3,341***	*314*	***$13,636***	***$9,147***	***-3.0***	***0.00***	***8.5***	***0.00***

6.4. Summary of Cost Evaluation Findings

Cost analyses estimated costs of intake, warning hearings, staffing meetings, office visits, drug tests, violation hearnings, arrests, state and county corrections, and residential treatment. Results were estimated for study participants for whom we had at least 6 months, 12 months, and 24 months of followup, allowing us to include more subjects for whom we had shorter follow-up periods and fewer subjects for whom we had longer follow-up. ***Six-month median costs were significantly higher for HOPE than PAU overall and in four sites; mean costs were higher overall and in three sites***. Overall sites, the mean cost for a HOPE probationer was $2,908 compared to $1,978 for a PAU probationer; median costs were also higher for HOPE--$1,698 compared to $364. There was considerable variability across groups and among sites—with median costs ranging from a low of $190 for PAU in Arkansas to a high of $2,723 for HOPE in Texas.

Results were similar for the 12-month sample, with median and mean costs significantly higher overall and in three sites. Mean (median) costs for the 12-month HOPE sample were $7,195 ($4,015) and for the 12-month PAU sample were $5,059 ($1,863). Again, we saw considerable variability, with median costs of $418 for PAU in Arkansas and $8,718 for HOPE in Texas.

For the 24-month sample, median and mean costs were significantly higher overall and in one site. Overall sites, mean (median) costs for the 24-month HOPE sample were $14,735 ($10,355) and for the

PAU sample were $11,413 ($5,797). Median costs ranged from $2,139 for PAU in Arkansas to $14,015 for HOPE in Oregon.

For each estimation period, cost differences were driven by treatment and incarceration costs.

7. Probation Experiences: Street, Treatment, Jail, and Prison Time

The evaluation collected detailed data on where participants were throughout the evaluation and the arrest and violation events that occurred. Specifically, we have detailed information on time spent in the community, jail, residential treatment, and prison; as well as the date of each arrest and violation. These data provide an opportunity to examine how probation was experienced by each of our eight study groups over the course of the evaluation—nearly 1,000 days for some.

Exhibit 7-1 shows where study participants were during their participation in the HOPE DFE by site and group. Each "bar" consists of a set of lines with each line representing one study participant. These graphs clearly demonstrate that the eight DFE groups had distinct experiences. The variation in the density of red lines clearly showing that HOPE cases experienced more jail (as we knew) than PAU. The prevalence of the green "street" time and absences of the blue "revoked" time in the Arkansas and Massachusetts PAU also visually demonstrates the relatively low revocation rates we observed for these two groups.

Exhibit 7-2 adds violation and arrest events to the graphic. The greater density of violations for the HOPE groups is easily visible—particularly for Oregon and Texas. Another view that shows only the first 100 days in the study for each participant is shown in ***Exhibit 7-3.*** Here, the jail stay (red line) following a violation (circle) or arrest (+) is clear for the HOPE groups in particular.

Exhibit 7-1. Time on the street, in residential treatment, in jail, or in prison/revoked for DFE participants

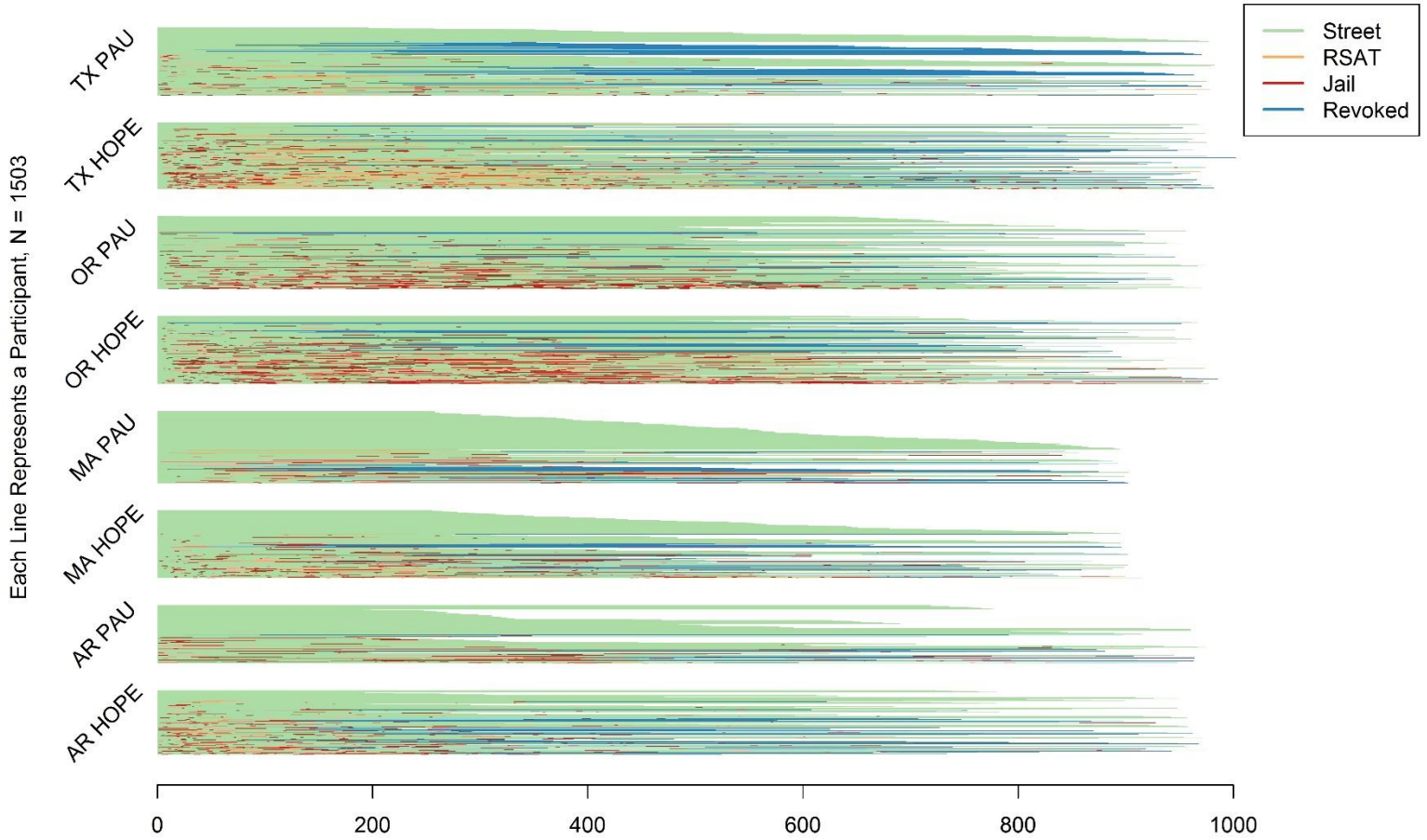


Exhibit 7-2. Time on the street, in residential treatment, in jail, or in prison/revoked for DFE participants with violations and arrests

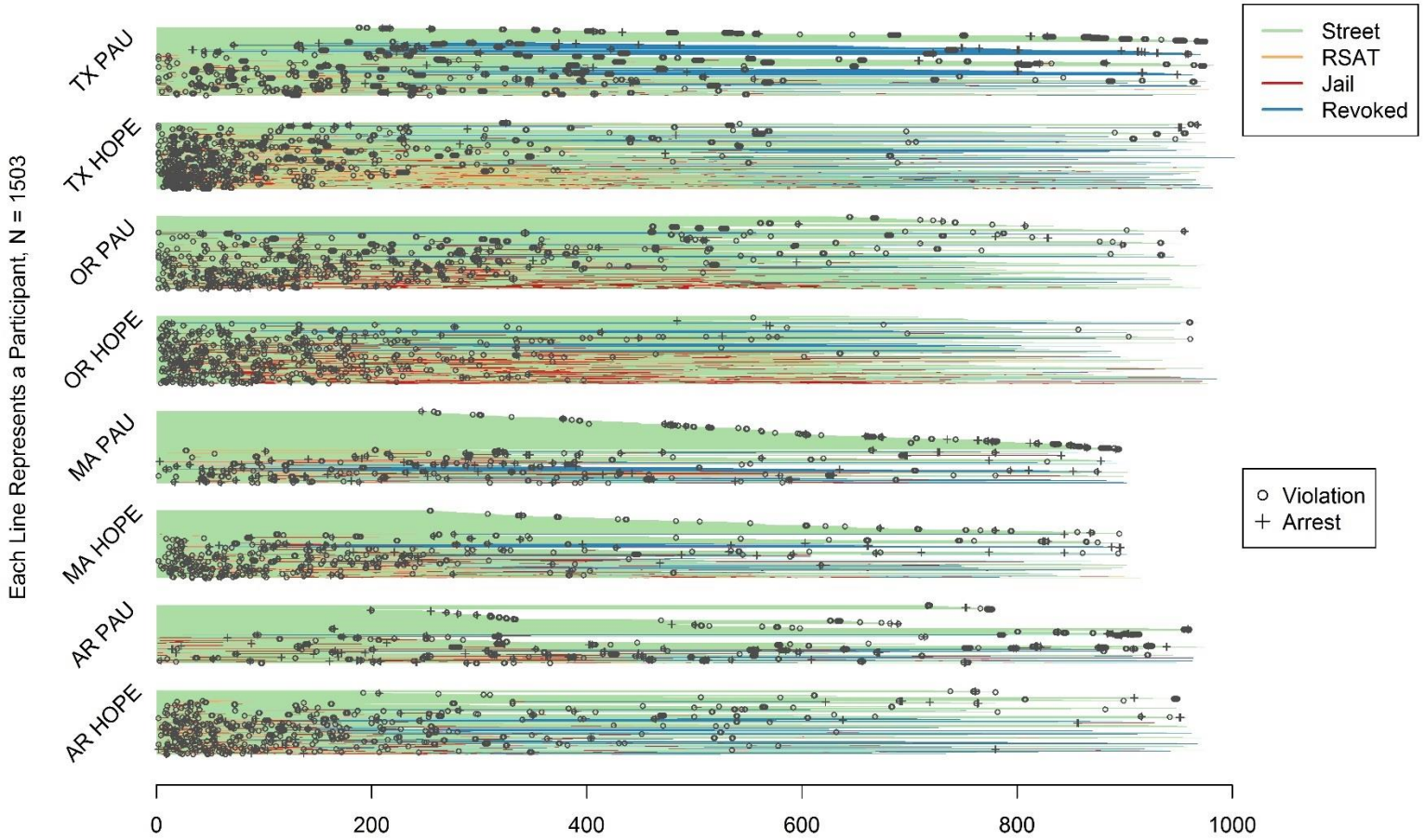
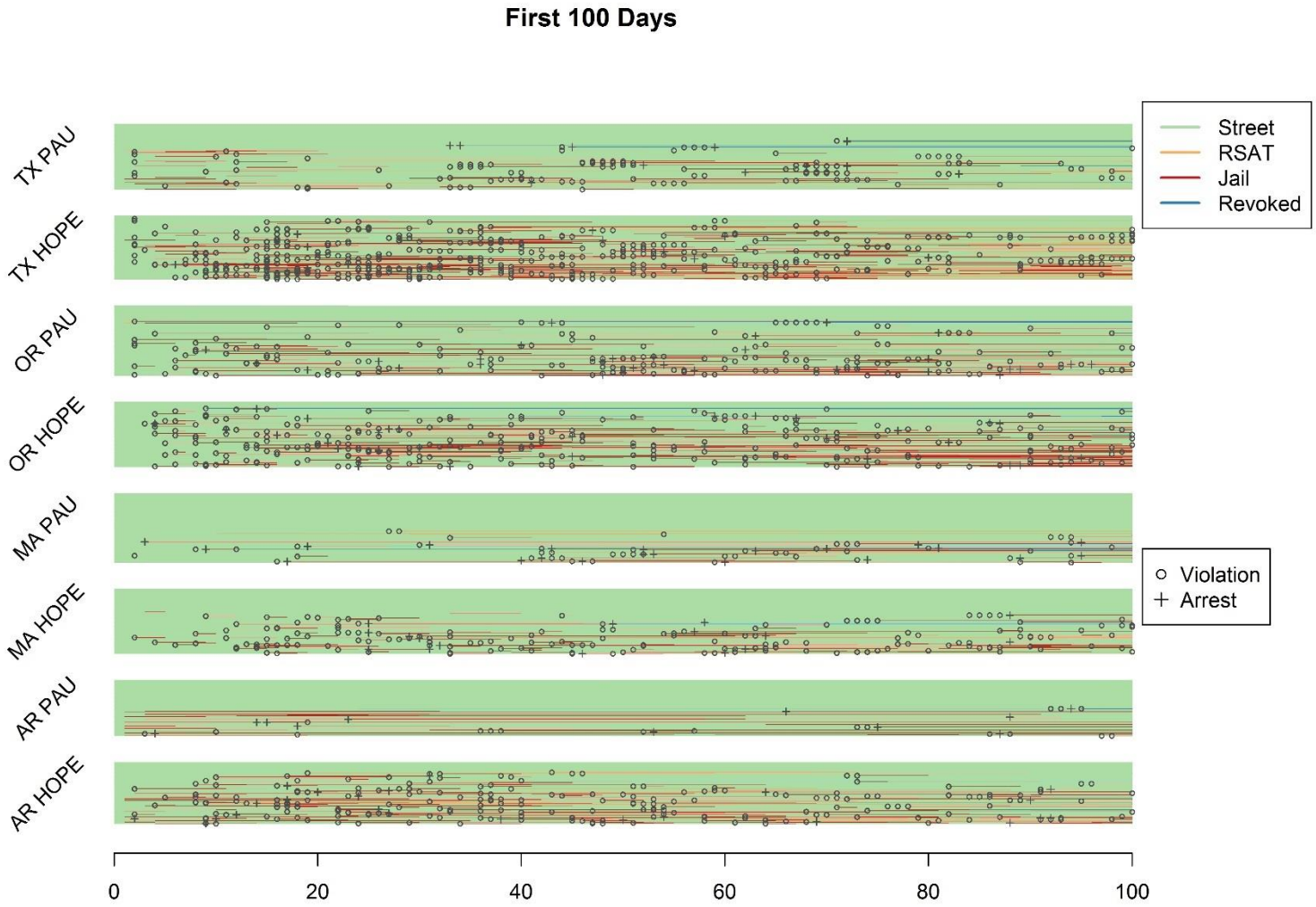
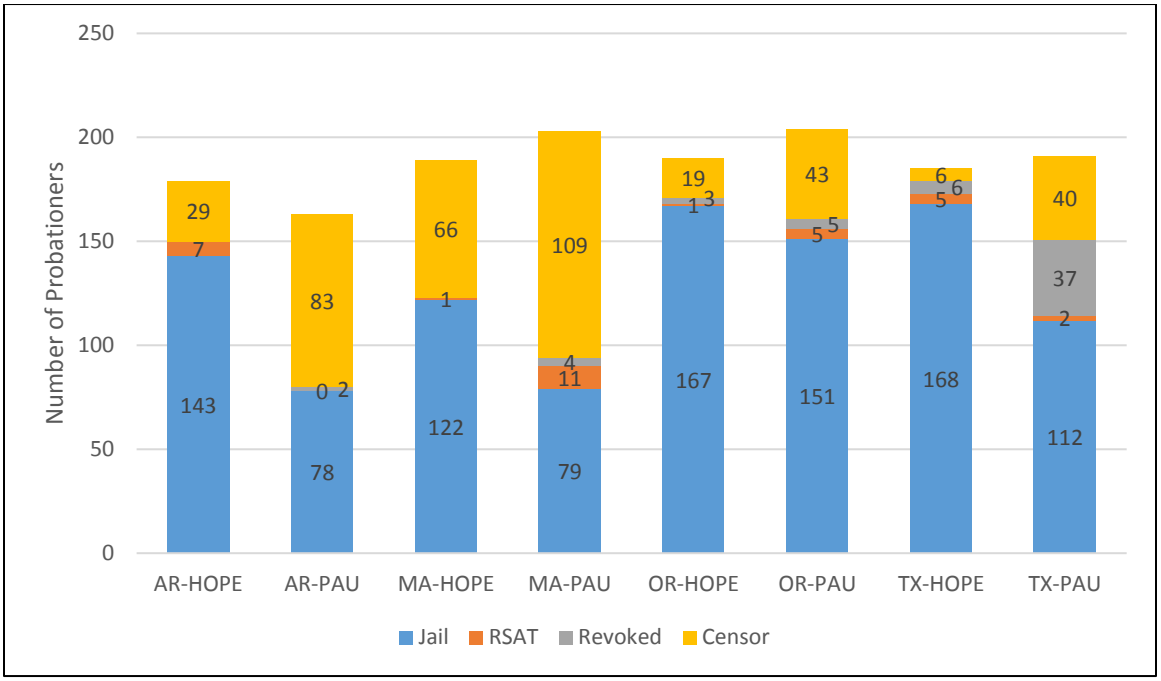


Exhibit 7-3. Time on the street, in residential treatment, in jail, or in prison/revoked for DFE participants, first 100 days



The first place-based transitions are shown by site and group in ***Exhibit 7-4***. Everyone began on the street, so these are street-to-next place[55].For HOPE participants in all four sites and PAU participants in two sites (Oregon and Texas), the most likely first transition was from the street to jail. In Arkansas and Massachusetts, PAU probationers were most likely to move directly to the study end ("Censor") without jail, residential treatment, or revocation. (This is apparent in the previous three exhibits by the dominance of green in the "bars" for these two groups.)

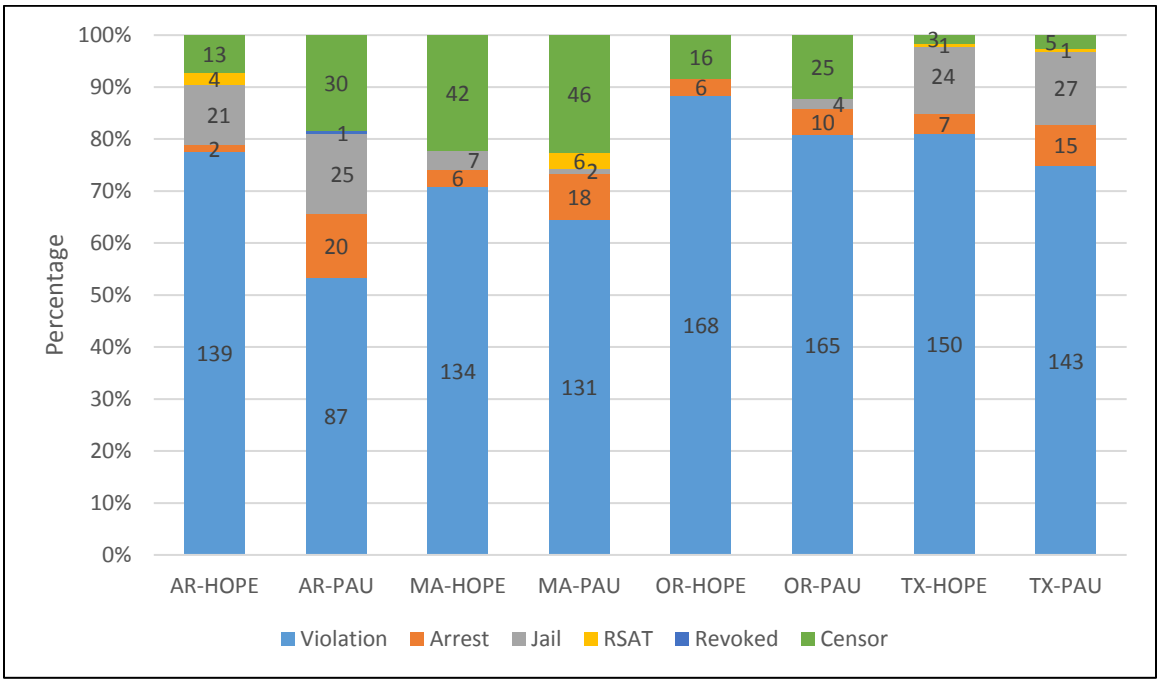
Exhibit 7-4. First place transition by site and group (street to ….)



If we include the events (violation and arrest) in addition to the places, more than 50% of all groups had a violation as the first experience (***Exhibit 7-5***). Arrests prior to a violation were rarer for the HOPE groups than the PAU groups. Notable numbers—particularly Arkansas PAU and Massachusetts probationers—transitioned out of the study without experiencing any event.

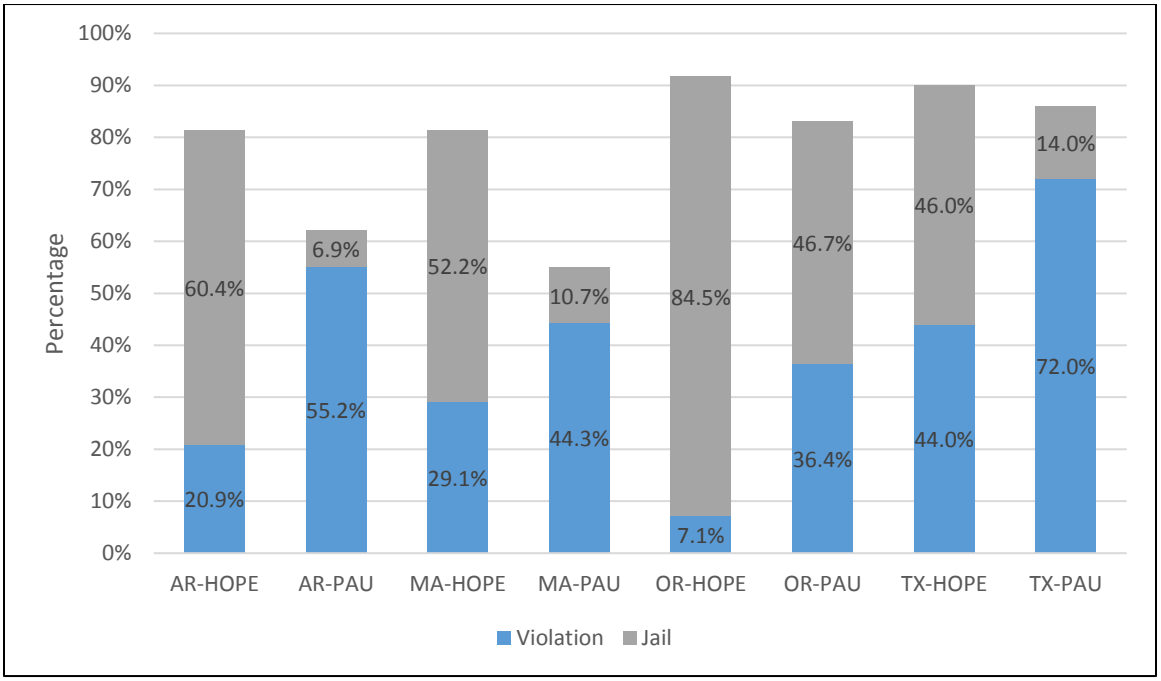
[55] A small handful of cases were in jail at the time of their random assignment to HOPE. There transitions were reset to begin with their initial release.

Exhibit 7-5. First place or event transition by site and group (street to ….)



HOPE was designed to assure that a violation was followed by a consequence. ***Exhibit 7.6*** shows the transition from an initial violation to another violation or jail (other outcomes not shown). HOPE probationers who violated were more likely to go to jail following an initial violation while PAU probationers were more likely to have a second violation without an intervening jail stay.

Exhibit 7-6. Transition from an initial violation to a second violation or to jail by site and group



The DFE study population experienced up to 70 transitions during the evaluation period. ***Exhibit 7-7*** and ***7-8*** shows all transitions summed for the HOPE and PAU groups[56]. (These transition summaries ignore the final transitions to 'censor' or end of study) The 743 HOPE probationers experienced 3,762 violations, 2,799 jail stays, 517 arrests, and 191 revocations. We again see that the most common response to a violation was jail—consistent with HOPE principles—2,217 violations were followed by jail with only 815 followed by another violation. In comparison, the 761 PAU probationers experienced 3,123 violations, 1,067 jail stays, 626 arrests, and 165 revocations—with sequential violations much more common that violation-jail transitions (1,541 versus 737).

Exhibit 7-7. HOPE probationer transitions summed

	Arrest	Died	Jail	Revoked	Res. Tx	Street	Violation
Arrest	73	0	197	26	3	45	83
Jail	43	1	73	70	78	2175	339
Revoked	49	0	0	0	0	0	7
Res. Tx	2	0	2	0	1	190	39
Street	107	5	310	33	124	0	2479
Violation	243	2	**2217**	62	44	249	**815**
Total	**517**	**8**	**2799**	**191**	**250**	**2659**	**3762**

Exhibit 7-8. PAU probationer transitions summed

	Arrest	Died	Jail	Revoked	Res. Tx	Street	Violation
Arrest	70	1	198	43	2	22	191
Jail	30	0	16	35	35	804	145
Revoked	35	0	0	0	0	0	11
Res. Tx	2	0	1	0	0	61	14
Street	141	2	115	9	28	0	1221
Violation	348	2	**737**	78	19	135	**1541**
Total	**626**	**5**	**1067**	**165**	**84**	**1022**	**3123**

Exhibit 7-9 shows the transitions summed by site for the HOPE probationers. The most common transition across the sites was the violation-to-jail transition. There were substantially more violations and jail stays in Oregon and Texas than in Arkansas and Massachusetts. The relatively large number of violation-to-violation transitions in Texas (478 compared to a hundred or so in the other sites) was due to a local policy that allowed probationers to turn themselves in if they missed a drug test or office visit;

[56] Each transition matrix captures the transition from one state to the next—e.g., from state 1 to state 2, from state 2 to state 3—the exhibits show the summation of the matrices across all transitions for each group. For example, the HOPE probationers experienced 73 sequential arrests across the study period and 2,217 violation-to-jail transitions (Exhibit 7-7).

if they failed to turn themselves in within 24 hours, they received violations for the missed test/visit and for failing to turn themselves in.[57]

Exhibit 7-9. HOPE probationer transitions summed by site

	Arrest	Died	Jail	Revoked	Res. Tx	Street	Violation
Arkansas							
Arrest	22	0	45	5	0	7	10
Jail	13	0	0	44	17	366	34
Revoked	19	0	0	0	0	0	0
Res. Tx	0	0	1	0	0	28	4
Street	18	1	46	3	8	0	462
Violation	48	0	**384**	7	8	30	**130**
Total	**120**	**1**	**476**	**59**	**33**	**431**	**640**
Massachusetts							
Arrest	25	0	63	8	0	14	20
Jail	13	0	3	22	18	328	31
Revoked	14	0	0	0	0	0	4
Res. Tx	1	0	1	0	1	12	17
Street	30	2	41	0	15	0	401
Violation	71	2	**312**	6	3	31	**134**
Total	**154**	**4**	**420**	**36**	**37**	**385**	**607**
Oregon							
Arrest	18	0	75	5	1	14	29
Jail	6	1	12	0	31	773	108
Revoked	3	0	0	0	0	0	2
Res. Tx	0	0	0	0	0	28	15
Street	35	2	48	5	9	0	861
Violation	95	0	**806**	22	5	69	**73**
Total	**157**	**3**	**941**	**32**	**46**	**884**	**1088**
Texas							
Arrest	8	0	14	8	2	10	24
Jail	11	0	58	4	12	708	166
Revoked	13	0	0	0	0	0	1
Res. Tx	1	0	0	0	0	122	3
Street	24	0	175	25	92	0	755
Violation	29	0	**715**	27	28	119	**478**
Total	**86**	**0**	**962**	**64**	**134**	**959**	**1427**

[57] In addition, at the beginning of the study, individuals were receiving a violation for failing to call the drug test hotline so someone who didn't call in and who missed a test received two violations. These violations were excluded from the analyses.

Exhibit 7-10 shows the transition summaries for the PAU probationers. Except for Oregon, most violations were followed by another violation (rather than jail). In Oregon, as with the HOPE probationers, violations were most frequently followed by a jail stay (481 times versus 356 times). These charts also show the differences in revocations across the sites and between groups—revealing (again) that relatively few PAU probationers (compared to HOPE) were revoked in Arkansas and Oregon.

Exhibit 7-10. PAU probationer transitions summed by site

	Arrest	Died	Jail	Revoked	Res. Tx	Street	Violation
Arkansas							
Arrest	29	1	49	4	0	5	55
Jail	8	0	0	7	2	94	19
Revoked	3	0	0	0	0	0	3
Res. Tx	0	0	0	0	0	2	0
Street	40	0	30	1	0	0	144
Violation	80	0	**49**	9	0	14	**211**
Total	**160**	**1**	**128**	**21**	**2**	**115**	**432**
Massachusetts							
Arrest	13	0	54	8	1	7	32
Jail	12	0	0	27	10	59	13
Revoked	8	0	0	0	0	0	3
Res. Tx	2	0	0	0	0	18	9
Street	26	1	8	3	11	0	180
Violation	80	2	**63**	8	10	16	**145**
Total	**141**	**3**	**125**	**46**	**32**	**100**	**382**
Oregon							
Arrest	11	0	85	6	1	9	60
Jail	9	0	15	0	7	503	82
Revoked	1	0	0	0	0	0	4
Res. Tx	0	0	1	0	0	18	3
Street	48	1	34	1	7	0	610
Violation	127	0	**481**	12	4	68	**356**
Total	**196**	**1**	**616**	**19**	**19**	**598**	**1115**
Texas							
Arrest	17	0	10	25	0	1	44
Jail	1	0	1	1	16	148	31
Revoked	23	0	0	0	0	0	1
Res. Tx	0	0	0	0	0	23	2
Street	27	0	43	4	10	0	287
Violation	61	0	**144**	49	5	37	**829**
Total	**129**	**0**	**198**	**79**	**31**	**209**	**1194**

Competing risk cause-specific Cox models were estimated by site for multiple transitions to examine the relationship of HOPE participation to transitions. ***Exhibit 7-11*** shows the results for the first transition between states—from the street to the next state (jail, prison/revoked, residential treatment) or event (arrest or violation). The last non-missing state was 'carried forward' to the next state in order to provide a 'censoring state'—for example, to include people who had no intervening events between intake and the end of the study in the analyses these individuals transitioned from street to street, time 1 to time 2. Entries are the coefficient estimates (standard errors) and [hazard ratios] for group effect on the time to state 2. (Competing states or events with less than 10 occurrences were dropped from the models.) The hazard ratio is the exponentiated coefficient. A negative coefficient yields a hazard ratio less than one and implies that a HOPE probationer has a lesser risk of failure per unit time (conditioned on not failing up to that time) than a PAU probationer. For example, the risk of arrest as the first event per unit time for a HOPE probationer in Arkansas is about 1/5th that of a PAU probationer.

Exhibit 7-11. Cause-specific Cox model results for group effect on time to state #2: coef, (se(coef)), [exp(coef)], and N

Site	Arrest	Violation	Jail	Revoked	Treatment	Street
Arkansas	**-1.615†** **(0.747)** **[0.199]** **22**	**1.142‡** **(0.140)** **[3.135]** **226**	-0.139 (0.298) [0.870] 46			0.143 (0.140) [1.153] 43
Massachusetts	-0.560 (0.476) [0.571] 24	**0.508‡** **(0.124)** **[1.662]** **265**				**0.565‡** **(0.219)** **[1.760]** **85**
Oregon	0.042 (0.521) [1.043] 16	**0.484‡** **(0.111)** **[1.622]** **333**				0.275 (0.331) [1.317] 40
Texas	0.299 (0.527) [1.349] 20	**1.098‡** **(0.121)** **[2.997]** **302**	-0.055 (0.304) [0.946 47]			

Note: Shaded cells indicate that the state occurred less than 10 times and estimates were not generated.
†HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.

The results for Arkansas indicate that the risk of arrest as the first event is less and the risk of a violation as the first event is greater for HOPE probationers than for PAU probationers. Results with respect to violations repeat for the other three sites—HOPE probationers are at greater risk of a violation than PAU probationers. In Massachusetts, HOPE probationers were more likely to time out of the study (street to street transition) without another intervening event than were PAU probationers.

Exhibit 7-12 shows the results for the second transition—from state 2 (whatever that may have been) to specific states at time 3. For all sites, HOPE probationers are at greater risk of jail at state 3 than PAU probationers and at lower risk of a violation as the third state.

Exhibit 7-12. Cause-specific Cox model results for group effect on time to state #3: coef, (se(coef)), [exp(coef)], and N

Site	Arrest	Violation	Jail	Revoked	Treatment	Street
Arkansas	**-0.674†** **(0.331)** **[0.509]** **40**	**-0.986‡** **(0.198)** **[0.373]** **114**	**1.956‡** **(0.298)** **[7.072]** **98**			0.524 (0.317) [1.689] 44
Massachusetts	**-1.153‡** **(0.332)** **[0.316]** **49**	**-0.629‡** **(0.172)** **[0.533]** **146**	**1.670‡** **(0.258)** **[5.310]** **93**			0.941 (0.598) [2.563] 15
Oregon	-0.619 (0.375) [0.539] 31	**-1.611‡** **(0.263)** **[0.200]** **92**	**0.997‡** **(0.143)** **[2.711]** **220**			
Texas	-0.788 (0.494) [0.455] 19	**-0.397‡** **(0.139)** **[0.672]** **214**	**1.268‡** **(0.264)** **[3.553]** **78**			**0.605†** **(0.281)** **[1.831]** **54**

Note: Shaded cells indicate that the state occurred less than 10 times and estimates were not generated.
†HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.

The next set of results (***Exhibit 7-13)*** show the results for the third transition—from state 3 to state 4. These results show that HOPE probationers are at much greater risk than PAU probationers to be transitioning to the street. This is directly linkable to the previous results which show that HOPE probationers were much more likely to have been in jail as state 3. PAU probationers across the sites are are much greater risk than HOPE probationers to experience a violation as state 4. This pattern is apparent for subsequent transitions (data not shown)—HOPE probationers are at greater risk of jail as state 5 and greater risk of returning to the street at state 6.

We estimated nine sets of these models—with the final set modeling the transitions from state 9 to state 10. (As noted, results are similar and are not provided here.) However, this last set is the first time we have sufficient transitions to residential treatment (10) to estimate the competing hazard for this state. Although results were not statistically significant (small sample size), the hazard ratio was 2.475 indicating a greater chance of treatment for HOPE cases than PAU cases (z = 1.272, p = 0.203).

Exhibit 7-13. Cause-specific Cox model results for group effect on time to state #4: coef, (se(coef)), [exp(coef)], and N

Site	Arrest	Violation	Jail	Revoked	Treatment	Street
Arkansas	**-0.941** **(0.436)** **[0.390]** **26**	**-0.522** **(0.205)** **[0.593]** **106**	0.284 (0.337) [1.329] 38			**0.949** **(0.266)** **[2.584]** **96**
Massachusetts	-0.524 (0.346) [0.592] 35	**-0.961** **(0.249)** **[0.382]** **75**	-0.234 (0.261) [0.791] 59			**0.998** **(0.289)** **[2.714]** **85**
Oregon	-0.451 (0.483) [0.637] 18	**-1.222** **(0.289)** **[0.295]** **68**	**-1.016** **(0.326)** **[0.362]** **48**			**0.653** **(0.151)** **[1.920]** **206**
Texas	-0.441 (0.518) [0.643] 20	**-0.644** **(0.158)** **[0.525]** **186**	**1.356** **(0.278)** **[3.881]** **74**	-1.125 (0.816) [0.325] 10		1.027 (0.279) [2.793] 69

Note: Shaded cells indicate that the state occurred less than 10 times and estimates were not generated.
†HOPE and PAU differ at $p < 0.05$; ‡HOPE and PAU differ at $p < 0.01$.

8. Limitations, Conclusions, and Recommendations for Policy and Practice

8.1. Limitations

The study has two primary limitations. First, we were unable to obtain arrest records from the FBI and had to rely on state and local data sources. Thus, our criminal recidivism indicators are limited to single states. Several of our sites were relatively close to state lines and individuals may have left their home states and been arrested elsewhere. We have no way of knowing whether and in what way HOPE probation may have been related to an individual probationer's decision to leave the state (likely not permissible under the terms of supervision). On the one hand, the closer supervision and requirements for regular contact for testing and so forth may mean that HOPE probationers would be less likely to leave. On the other hand, the conditions themselves could have encouraged some to leave (abscond). Absent any additional information, we assume that HOPE and PAU probationers were equally likely to leave their jurisdictions.

Second, the response rates for the 6- and 12-month follow-up interviews were disappointing. The study was designed around the on-site research coordinators serving as interviewers, beginning with study enrollment. These research coordinators were stationed in the probation offices in three sites and across the street from the court house in the fourth site so that they would be available on a full-time basis to do study intake and the baseline interview. The original design assumed that baseline identification of 400 HOPE-eligible probationers would be accomplished in 6 to 9 months—in other words, that intake would be (largely) completed about the time that the research coordinators needed to be able to go out into the community (or to the jail) to conduct follow-up interviews. As noted in the methods section, study enrollment continued for 17 months in one site, 22 months in one site, and 25 months in the other two sites. Data collection continued for 6 months following the final intake. The impact of this overlap in enrollment and follow-up was to hinder the ability of the research coordinators to go into the field and do follow-up interviews. We have criminal justice and drug test results for all study participants, so the primary loss was information from more individuals on other outcomes (e.g., employment) and on attitudes. Further, as noted in the methods section, response bias analyses found few differences between our respondents and non-respondents.

8.2. Discussion

Four sites that differed in organizational structures and populations successfully implemented HOPE programs—holding probationers accountable to their conditions of supervision and reducing drug use. HOPE was effective in increasing compliance with some supervision conditions (e.g., probation officer visits and payment of fees and fines). HOPE probation also appeared to have positive effects on drug use based on oral swab drug tests conducted in conjunction with interviews 6 and 12 months after program enrollment. HOPE probationers were randomly tested at a very high rate (26,991 tests compared to 4,942 for the PAU probationers). This extensive testing led to more violations—as would have been expected—although positive tests reduced substantially over time (again suggesting positive impacts on drug use).

HOPE probationers were more likely to go to jail (82% versus 56%), to have more jail stays (3.8 versus 1.4), and to spend more days in total in jail (47 versus 33 days). HOPE probationers were also more likely to be sent to residential treatment (overall and in three sites).

Overall, HOPE did not reduce recidivism, as measured by arrest, revocation, and new conviction. In two sites, revocations were higher for HOPE than PAU and in one site reconvictions were higher. The sole signficant positive recidivism finding was a longer time to revocation in one site (although final revocation rates were similar).

More jail days, more residential treatment, and similar (or higher) recidivism resulted in higher (although not always significantly higher) costs for HOPE compared with PAU.

In addition to program fidelity, it seems that HOPE probationers understood what was expected of them. The HOPE and PAU probationers we interviewed reported a clear understanding of the terms of their probation. Both study groups had a strong sense that their probation officer would find out about any noncompliance and would arrest them or have them arrested for noncompliance. Both groups also had a strong sense that the judge would do something in response to noncompliance, although HOPE probationers at their 12-month interview were more certain than PAU probationers that the judge would respond suggesting that the HOPE probationers—if they didn't understand initially—learned that sanctions would happen. Through qualitative interviews with HOPE probationers we learned that they were universally aware of the specific terms and conditions that HOPE imposed upon them and what was required of them under HOPE. Most had at least one violation and reported that all violations were met with some sort of consequence, typically a brief jail stay. ACASI interviews with probationers show similar findings, with HOPE probationers significantly more likely than PAU probationers to report spending time in jail, including receiving jail time because of a positive drug test.

HOPE probationers who participated in a qualitative interview were mixed as to whether they thought about the consequences before committing violations. Most did not report giving the possibility of punishment much thought (even though they knew in an intellectual sense what *could* happen), with some not actually caring much about being punished. For some, though, the deterrence message set in over time, leading them to be more thoughtful about their behavior. ACASI interviews with probationers underscore this point: HOPE probationers were more sensitive to the possible consequences of noncompliance (as measured by the deterrence score) and reported a lower tolerance for law violations than their PAU counterparts.

The ACASI interviews also offer some evidence about change in other attitudes among HOPE probationers. Specifically, HOPE probationers reported greater self-efficacy and a lower level of identification with crime-involved people than PAU probationers. Despite these positive changes, HOPE probationers reported a lower level of substance abuse treatment motivation than PAU probationers.

ACASI interviews—and our extensive transition analyses—show that HOPE and PAU probationers experienced probation differently. Although HOPE and PAU probationers were equally likely to be required to attend substance abuse treatment as a condition of supervision, HOPE probationers were more likely to attend treatment. More HOPE probationers than PAU probationers were subjected to drug testing as a supervision requirement and very few PAU probationers were subjected to random testing. HOPE probationers who participated in a qualitative interview felt that the most difficult part of HOPE was balancing the need to report for frequent drug tests with their work schedules, leading some to lose jobs due to their participation in HOPE. They also felt that the emphasis on accountability in

HOPE was helpful, as was the structure that it provided to their daily lives, which was often lacking before HOPE.

ACASI interviews with HOPE and PAU probationer violators show that the majority in each study group felt that the violation that they received was not a surprise, and the punishment they received for violating probation was fair. Qualitative interviews with HOPE probationers show that they generally thought that HOPE was a fair program and that it was up to them to make something out of it.

Although our interviews with HOPE probationers were limited in scope and depth, and may not be representative of the overall HOPE probationer experience within this DFE, what we learned from them and from the exhaustive interviews with the HOPE teams leads to some preliminary observations about who might benefit the most from HOPE. We posit a sort of curvilinear relationship between offender type and HOPE effectiveness that suggests two types of offenders for whom HOPE *may* be more beneficial.

First are what we would term the early career probationers—individuals who have not been engaged for long in the criminal lifestyle, who have limited exposure to incarceration (and indeed, some of the probationers we interviewed had never been in custody) and who were highly motivated to stay out of custody and get on with their lives. Some reported that they did not like HOPE and did not see HOPE as fair, but indicated that HOPE provided them a chance to change for the better. These were often people who needed help with basic life skills, and HOPE helped them manage their lives more successfully. For these probationers, HOPE could be a time out, allowing them to reset and correct course.

Second, at the other end of the spectrum, are those who could be termed "late career probationers." These individuals had been in and out of trouble (and custody) for many years, were often older (but not always), and were looking for a way out of a criminal lifestyle. For them, HOPE provided a chance to redeem themselves, to avoid "dying in prison," and to "go straight" or "make good" while there was still time (Maruna, 2001). One mid 50's HOPE probationer had been a drifter most of his life, was living in a homeless camp and had spent, by his own account, about half of his life in custody. He spoke to us of being tired of the grind and wanting to make something of himself before he died. He told us his goal was to be "the poster child" for HOPE. And, according to the HOPE team at that site, he was. He was usually there an hour early for his drug test, never missed a meeting and at the time of our interview was nearing completion of his probation term in HOPE and had no violations. He was sober and employed (albeit marginally). For him, these outcomes signified progress—progress he attributed to HOPE. Thus, these late career HOPErs were burned out from their involvement in the criminal justice system and saw HOPE as one final chance to change. Again, they were not always happy with HOPE, but were responsive to it. Or, perhaps these were offenders who were aging out of crime and would have desisted even without HOPE (or any intervention).

In between these two groups were the mid-career probationers. These individuals had accumulated quite a bit of experience with the criminal justice system, were highly antisocial and embedded in criminal networks, and were quite happy with their status quo. They were not troubled by the prospect of going to prison (because they had been there repeatedly, but not enough to be weary of it), were more concerned with present gratifications than with future consequences, often had little of substance going in their lives that could actually be disrupted by a short and immediate jail stay under HOPE, and were dismissive of efforts to persuade them to change their behavior. HOPE to many of them was a joke

(e.g. “it’s easy to fool the drug testing system”), something they had to put up with (because under the DFE they had no choice but to be in HOPE) until they completed their terms and could go back to what they were doing before. Even if they thought HOPE was fair, they thought it inconsequential for their lives and reflected their deeper set of criminal cognitions and decision making states. These individuals were too far along to be intercepted (as was the case for the early career probationers), but not yet burned out from long experience with the system (as was the case for the late career probationers).

The identification of these three types of individuals relates to the concept of *deterability* and the context of deterrence in a case. Specifically, it may signify the importance of understanding the ability of the individual offender to be reached by a deterrence message (Jacobs, 2010).

8.3. Conclusions and Implications for Policy and Practice

HOPE probation has been widely promoted and adapted as a means for substantially improving probation outcomes while generating cost savings[58]. The findings of this rigorous four-site randomized controlled trial suggest otherwise. So what is to be made of this?

The results do not say do not implement HOPE or similar programs based on “swift, certain, and fair” principles. The results suggest that great consideration should be given to the implications of HOPE programs within the context of current probation policy and practice—PAU context is important. Within the DFE:

- PAU revocation rates were low (9% and 13%) in two sites—suggesting limited ability to reduce revocations and that sites with low revocation rates should consider whether to implement procedures to mitigate any potential increases in revocations that would accompany the increased surveillance of HOPE.
- In at least two sites, revocation could yield only short prison stays (90 days)—suggesting limited opportunities for “prison bed savings” even if revocations were lower with HOPE and providing a smaller incentive for individuals to comply.
- PAU was based at least somewhat on Risk-Needs-Response principles in at least two sites—suggesting an additional consideration with respect to the integration of HOPE with PAU.
- In one site, probation could use short jail stays on their authority (and did for PAU cases)—suggesting that a HOPE judge was not necessary to enforce conditions.

Thus, the similar outcomes may hinge on the “compared to what” aspect of any evalution—in that findings suggest that HOPE worked as well as but not better than PAU. However, given the consistency of findings across four sites that differed in the administration of PAU, there is little to support a conclusion that HOPE or HOPE-like programs will produce substantial improvements over PAU when implemented widely. This may be particularly true given the extensive resources that were devoted to facilitating implementation in the DFE—resources likely much greater than sites would generally have to implement a new program.

[58] As noted earlier, Judge Alm now claims that HOPE was always predicated on an RNR/MI approach to probation. This is not how the model was prescribed by BJA (or others) at the time the DFE was initiated. This evaluation tested the model as the surveillance/deterrence model that was promoted at the time the DFE was initiated and HOPE was being widely adopted as a “swift-certain-fair” probation model with treatment reserved for those unable to stop drug use to comply with a random testing regimen.

Additional research is needed to determine whether there may be specific types of probationers who would be more responsive to this type of program. For example, as discussed above, perhaps there are subpopulations for whom the threat of even short jail stays may provide more motivation. Perhaps HOPE is most suitable for offenders who need some intervention but who are not so deeply entrenched in a criminal lifestyle that the threat of sanctioning is insufficient to elicit behavioral change. The structure of HOPE may also prove useful in helping individuals who want to change to stay away from drugs and better manage their lives. At least anecdotally it also may be that HOPE provides a useful crutch for those who have decided to leave drugs and a criminal lifestyle behind. We heard from multiple individuals that there were HOPE probationers who wanted to remain on the random drug testing regime because it was helping them stay sober. More research is needed to develop our understanding of who (if anyone) is likely to maximally benefit from a swift and certain sanctioning regimen such as HOPE.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716–723.

Akers, R. L. (1990). Rational choice, deterrence, and social learning theory in criminology: The path not taken. *Journal of Criminal Law and Criminology*, 81(3), 653–676.

Alm, S. S. (2015). HOPE probation and the new drug court: A powerful combination. *Minnesota Law Review,* 99, 1665–1695.

Alm, Steven S. 2016. HOPE probation: Fair sanctions, evidence-based principles, and therapeutic alliances. *Criminology & Public Policy*, 15(4), 1195–1214. doi:10.1111/1745-9133.12261.

Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct*. Cincinnati, OH: Anderson.

Andrews, D. A., & Dowden, C. (2005). Managing correctional treatment for reduced recidivism: A meta-analytic review of programme integrity. *Legal and Criminological Psychology,* 10(2), 173–187.

Apel, R. (2013). Sanctions, perceptions, and crime: Implications for criminal deterrence. *Journal of Quantitative Criminology*, 29(1), 67–101.

Apel, R., & Nagin, D. S. (2011). Deterrence. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*.

Argyris, C. (1985). *Strategies, Change and Defensive Routines*. Boston, MA: Pittman.

Association of Arkansas Counties (2011). Local Government Inmate Cost Report. Retrieved September 6, 2016, from http://www.arkleg.state.ar.us/assembly/2013/Meeting%20Attachments/000/I10248/Exhibit%20H.7a%20-%20Association%20of%20Counties%20-%20Local%20Government%20Inmate%20Cost%20Report.pdf

Baker, C. L., Bennet, D., & O'Brien, C. H. (2015). Quarterly Report on the Status of Prison Overcrowding, First Quarter 2015 (Report No. 15–104-DOC-01). Retrieved September 6, 2016, from http://www.mass.gov/eopss/docs/doc/research-reports/overcrowding/2015-1stqtr-overcrowding-report.pdf.

Beccaria, C. (1963). *On crimes and punishment* (H. Paolucci, Trans.). Indianapolis, IN: Bobbs-Merrill. (Original work published 1764)

Benichou, J., & Gail, M. H. (1990). Estimates of absolute cause-specific risk in cohort studies. Biometrics, 813-826.

Bentham, J. (1789). *An introduction to the principles of morals and legislation*. Oxford, UK: Oxford University Press.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy,* 76, (2), 169–217.

Block, M. K., & Heineke, J. M. (1975). A labor theoretic analysis of the criminal choice. *American Economic Review,* 65, 314–325.

Bucklen, K. B. & Zajac, G. (2009). But some of them don't come back (to prison!): Resource deprivation and thinking errors as determinants of parole success and failure. *The Prison Journal*, 89(3), 239–264.

Bunten, A., Kendall-Taylor, N., & Lindland, E. (2011). *Caning, context, and class: Mapping the gaps between expert and public understandings of public safety*. Washington, DC: Frameworks Institute.

Buntin, J. (2009, October 31). Swift and certain: Hawaii's probation experiment. *Governing.* Retrieved from http://www.governing.com/topics/public-justice-safety/Swift-and-Certain-Hawaiis.html

Bureau of Justice Statistics (2012). Justice Expenditure and Employment Extracts Program [Data file].

Bureau of Justice Statistics (2015). Inmates in custody of state or federal correctional facilities, including private prison facilities, December 31, 1999–2014 [Data file].

Bureau of Labor Statistics (2014). Occupational Employment Statistics. Retrieved January 27, 2016, from http://www.bls.gov/oes.

Butts, C. (2015). Network: The Statnet Project (http://www.statnet.org). R package version 1.12.0, http://CRAN.R-project.org/package=network.

Carmichael, S., & Piquero, A. R. (2004). Sanctions, perceived anger, and criminal offending. *Journal of Quantitative Criminology*, 20, 371–393.

Carns, T. W. & Martin, S. (2011). *Anchorage PACE: Probation Accountability with Certain Enforcement: A Preliminary Evaluation of the Anchorage Pilot PACE Project.* Anchorage: Alaska Judicial Council. Retrieved from http://www.ajc.state.ak.us/reports/pace2011.pdf.

Clear, T.R., & Frost, N.A. (2016*). The Punishment Imperative: The Rise and Failure of Mass Incarceration in America*. New York: New York University Press.

Cohen, M. A. (1988). Pain, suffering, and jury awards: A study of the cost of crime to victims. *Law and Society Review*, 22(3), 537–555.

Cook, P.J. (2016). Behavioral science critique of HOPE. *Criminology and Public Policy 15,4: 1155-1161.*

Cornish, D. B., & Clarke, R. V. (1986). The reasoning criminal: Rational choice perspectives on offending. New York: Springer-Verlag.

Cullen, F.T., Manchak, S.M., &Duriez, S.A. (2014). Before adopting Project HOPE, read the warning label: A rejoinder to Kleiman, Kilmer, and Fisher's comment. *Federal Probation*, 78(2), 75-77.

Cullen, F. T., Pratt, Travis C., & Turnanovic, Jillian J. (2016). It's hopeless: Beyond zero-tolerance supervision. *Criminology & Public Policy*, 15(4), 1215–1239.

Cullen, F.T., Pratt, T.C., Turanovic, J.J., and Butler, L. (2018). When bad news arrives: Project HOPE in a post-factual world. Journal of Contemporary Criminal Justice, 34(1), 13-34.

DeMichele, M., & Payne, B. K. (2012). Measuring community corrections officials' perceptions of goals, strategies, and workloads from a system perspective: Differences between directors and nondirectors. *Criminology & Penology*, 49(3), 388–410.

DuFrene, B. A., Noell, G. H., Gilbertson, D. N., & Duhon, G. J. 2005. Monitoring Implementation of Reciprocal Peer Tutoring: Identifying and Intervening with Students Who Do Not Maintain Accurate Implementation. *School Psychology Review,* 34(1), 74–86.

Duriez, S. A., Cullen, F. T., & Manchak, S. M. (2014). Is Project HOPE creating a false sense of hope: A case study in correctional popularity. *Federal Probation*, 78, 57–60.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. American *Journal of Community Psychology,* 41(3/4), 327–350.

Durlauf, S. N., & Nagin, D. S. (2011). Imprisonment and crime: Can both be reduced? *Criminology and Public Policy*, 10(1), 13–54.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18(2), 237–256.

Ehrlich, I. (1973). Participating in illegitimate activities: A theoretical and empirical investigation. *Journal of Political Economy,* 81,521–565.

Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science,* 5(1), 47–53.

Esbensen, F.-A., Matsuda, K. N., Taylor, T. J., & Peterson, D. (2011). Multimethod strategy for assessing program fidelity: The national evaluation of the revised G.R.E.A.T. program." *Evaluation Review*, 35(1), 14–39.

Essex Sheriff's Department. (2015). Historical Spending Categories. Retrieved September 6, 2016, from http://www.mass.gov/bb/gaa/fy2016/app_16/dpt_16/hosde.htm.

Fagan, A. (2013). Family focused interventions to prevent juvenile delinquency: a case where science and policy can find common ground. *Criminology & Public Policy*, 12(4), 617–650.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Gandhi, A. G., Murphy-Graham, E., Petrosino, A., & Schwartz Chrismer, S. (2007). The devil in the details: Examining the evidence for "proven" school-based drug abuse prevention programs. *Evaluation Review*, 31(1), 43–74.

Gendreau, P., Smith, P., & Theriault, Y. L. (2009). Chaos theory and correctional treatment: Common sense, correctional quackery and the law of fartcatchers. *Journal of Contemporary Criminal Justice, 25(4), 384–396*.

Gerds, T. A., & Scheike, T. H. (2015). *riskRegression: Risk Regression Models for Survival Analysis with Competing Risks*. R package version 1.1.7. https://CRAN.R-project.org/package=riskRegression

Gottfredson, D. C. & Gottfredson, G. D. (2002). Quality of school-based prevention programs: results from a national survey. *Journal of Research in Crime and Delinquency*, 39, 3–35.

Greenwood, C. R., Tapia, Y., Abbott, M., & Walton, C. (2003). A building-based case study of evidence-based literacy practices: Implementation, reading behavior and growth in reading fluency, K–4." *The Journal of Special Education,* 37(2), 95–110.

Hamilton, Z., Campbell, C. M., van Wormer, J., Kigerl, A. and Posey, B. (2016). Impact of swift and certain sanctions. *Criminology & Public Policy*, 15: 1009–1072. doi:10.1111/1745-9133.12245

Hamilton, Z., van Wormer, J., Kigerl, A., Campbell, C. M., & Posey, B. (2015). *Evaluation of Washington State Department of Corrections (WADOC) Swift and Certain (SAC) Policy Process, Outcome and Cost-Benefit Evaluation – Final Report.* Spokane: Washington State University, Institute for Criminal Justice. Retrieved from https://wsicj.wsu.edu/wp-content/uploads/sites/436/2015/11/SAC-Final-Report_2015-08-31.pdf.

Handcock M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2003). ERGM: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. Statnet Project, Seattle, WA. Version 2, http://www.statnet.org.

Harrell, A., & Roman, J. (2001). Reducing drug use and crime among offenders: The impact of graduated sanctions. *Journal of Drug Issues*, 31, 207–231.

Hartney, C. & Marchionna, S.. (2009). *Attitudes of US voters toward nonserious offenders and alternatives to incarceration.* Retrieved from http://www.nccdglobal.org/sites/default/files/publication_pdf/focus-voter-attitudes.pdf.

Hatry, H., Winnie, R., & Fisk, D. (1981). *Practical program evaluation for state and local governments*. *2nd Edition.* Washington, D.C.: The Urban Institute.

Hawken, A., & Kleiman, M. (2007, April 10). HOPE for reform: What a novel probation program in Hawaii might teach other states. *The American Prospect.* Retrieved from http://prospect.org/cs/articles?articleId=12628

Hawken, A., & Kleiman, M. (2009). *Managing drug involved probationers with swift and certain sanctions: Evaluating Hawaii's HOPE*. Washington, DC: U.S. Department of Justice, National Institute of Justice.

Hawken, A., Kulick, J., Smith, K., Mei, J., Zhang, Y., Jarman, S., Yu, T., Carson, C., & Vial, T. (2016). *HOPE II: A Follow-up to Hawai'i's HOPE Evaluation*. (Report # 2010-IJ-CX-0016). Washington, DC: U.S. Department of Justice, National Institute of Justice.

Jacobs, B. A. (2010). Deterrence and deterrability. *Criminology*, 48(2), 417–441.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica,* 47(2), 263–291.

Kahneman, D., & Tversky, A. (1982). On the psychology of prediction. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press (pp. 32–47).

Kennedy, D. (1997). Pulling levers: Chronic offenders, high-crime settings, and a theory of prevention. *Valparaiso University Law Review,* 31, 449–484.

Krisberg, B., & Marchionna, S. (2006). Attitudes of U.S. voters toward prisoner rehabilitation and reentry policies. FOCUS: Views from the National Council on Crime and Delinquency. Retrieved April 6, 2017 from http://research.policyarchive.org/5943.pdf.

Krivitsky, P. N. (2012). Exponential-Family Random Graph Models for Valued Networks. *Electronic Journal of Statistics*, 6, 1100–1128. Doi:10.1214/12-EJS696

Lake, C., Gotoff, D., & Pultorak, K. (2013). *Reducing incarceration levels in the U.S.: Opportunities for reform*. Open Society Foundation.

Latessa, E. J., Cullen, F. T., & Gendreau, P. (2002). Beyond Correctional Quackery– Professionalism and the Possibility of Effective Treatment. *Federal Probation,* 66(2), 43–49.

Latessa, E. J., Smith, P., Lemke, R., Makarios, M. D., & Lowenkamp, C. (2009). *Creation and Validation of the Ohio Risk Assessment System: Final Report*. Prepared for the Ohio Department of Rehabilitation of Correction.

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior and Organization,* 17, 377–400.

Lattimore, P. K., Krebs, C. P., Koetse, W., Lindquist, C. H., & Cowell, A. J. (2005). Predicting the effect of substance abuse treatment on probationer recidivism. *Journal of Experimental Criminology,* 1, 159–189.

Lattimore, P. K., MacKenzie, D. L., Zajac, G., Dawes, D., Arsenault, E., & Tueller, S. J. (2016). Outcome findings from the HOPE demonstration field experiment. *Criminology and Public Policy,* 15, 4, 1103–1141*,* doi:10.1111/1745-9133.12248.

Lattimore P. K., & Visher, C. A. (2009). *Multi-site evaluation of SVORI: Summary and synthesis.* Raleigh, NC: RTI International.

Lattimore, P. K., & Visher, C. A. (nd). Serious and Violent Offender Reentry Initiative (SVORI) Multi-site Impact Evaluation, 2004-2011 [United States]. ICPSR27101-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-11-07. http://doi.org/10.3886/ICPSR27101.v1

Lattimore, P. K., & Witte, A. D. (1986). Models of decision making under uncertainty: The criminal choice. In R. Clarke & D. Cornish (Eds.), *The reasoning criminal* (pp. 129–155). New York: Springer-Verlag.

Laven, D. N., Krymkowski, D. H., Ventriss, C. L., Manning, R. E., & Mitchell, N. J. (2010). From partnerships to networks: New approaches for measuring US National Heritage Area effectiveness. *Evaluation Review*, 34(4), 271–298.

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders*, 4, 124–147.

Lowenkamp, C. T., Latessa, E. J., & Smith, P. (2006). Does correctional program quality really matter? The impact of adhering to the principles of effective intervention. *Criminology and Public Policy,* 5, 575–594.

Lubbers, M. J., & Snijders, T. A. (2007). A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social Networks*, 29(4), 489–507.

MacKenzie, D. L. (2002). Reducing the criminal activities of known offenders and delinquents: crime prevention in the courts and corrections (from *Evidence-Based Crime Prevention*, P 330-404, 2002, Lawrence W. Sherman, David P. Farrington, et al, eds.)

MacKenzie, D. L. (2006). *What Works in Corrections*. New York, NY: Cambridge University Press.

MacKenzie, D. L., & Zajac, G. (2013). What Works in Corrections: The Impact of Correctional Interventions on Recidivism. Monograph commissioned by The National Academy of Sciences Committee on the Causes and Consequences of High Rates of Incarceration.

MacKenzie, D. L., Browning, K., Skroban, S., & Smith, D. (1999). The impact of probation on the criminal activities of offenders. *Journal of Research in Crime and Delinquency,* 36, 423–453.

Maruna, S. (2001). *Making good: How ex-convicts reform and rebuild their lives.* Washington, DC: American Psychological Association.

Matsueda, R.L., Kreager, D.A., & Huizinga, D. (2006). Deterring delinquents: A rational choice model of theft and violence. *American Sociological Review*, 71, 95–122. Retrieved from http://users.soc.umn.edu/~uggen/Matsueda_ASR_06.pdf.

Mihalic, S., & Irwin, K. (2003). From research to real world settings: Factors influencing the successful replication of model programs. *Youth Violence and Juvenile Justice*, 1, 307–329.

Moely, B. E., McFarland, M., Miron, D., Mercer, S., & Ilustre, V. (2002). Changes in college students' attitudes and intentions for civic involvement as a function of service-learning experiences. *Michigan Journal of Community Service Learning*, 9(1), 18–26

Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. *Journal of Statistical Software***,** 24(4), 1548–7660. http://www.jstatsoft.org/v24/i04/.

Nagin, D. S. (2007). Moving choice to center stage in criminological research and theory: The American Society of Criminology 2006 Sutherland Address. *Criminology,* 45, 259–272.

Nagin, D. S. (2013). Deterrence: A review of the evidence by a criminologist for economists. *Annual Review of Economics,* 5, 83–105.

Nagin, D. S., & Paternoster, R. (1993). Enduring individual differences and rational choice theories of crime. *Law and Society Review,* 27, 467–496.

Nice, M. (2002). *Oregon's County Jails: A Brief Comparative Analysis* (Report No. 001-02). Retrieved September 6, 2016, from https://multco.us/file/26863/download.

Norman, R. A. (2012). *Inmate Cost Report - 2010: Comparison of Cost Per Day Reported and Verified*. Retrieved September 6, 2016, from http://www.arkleg.state.ar.us/assembly/2013/Meeting%20Attachments/000/I10248/Exhibit%20H.7b%20-%20DLA%20-%20Local%20Inmate%20Cost%20report%20verification.pdf.

O'Connell, M. E., Boat, T., & Warner, K. E. (Eds.). (1994). *Preventing mental, emotional and behavioral disorders among young people: Progress and possibilities.* Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Youth and Young Adults: Research Advances and Promising Interventions. Board on Children, Youth and Families. Division of Behavioral and Social Sciences and Education. National Research Council and Institute of Medicine of The National Academies. Washington, DC.

O'Connell, D. J., Brent, J. J., & Visher, C. A. (2016). Decide your time. *Criminology & Public Policy*, 15, 1073–1102.

O'Connell, D. J., Visher, C. A., Martin, S., Parker, L., & Brent, J. J. (2011). Decide your time: Testing deterrence theory's certainty and celerity effects on substance-using probationers. *Journal of Criminal Justice,* 39, 261–267.

Olds, D. L. (2002). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science,* 3, 153–172.

Olds, D. L., Hill, P. L., O'Brian, R., Racine, D., & Moritz, P. (2003). Taking preventative intervention to scale: The nurse-family partnership. *Cognitive and Behavioral Practice,* 10(4), 278–290.

Oleson, J. C. (2016). HOPE springs eternal: New evaluations of correctional deterrence. *Criminology & Public Policy*, 15, 1163–1184.

Paternoster, R. (1987). The deterrent effect of the perceived certainty and severity of punishment: A review of evidence and issues. *Justice Quarterly,* 4, 173–217.

Pearlin, L., & Schooler, C. (1978). The structure of coping. *J. of Health and Social Behavior*, 19, 2-21. http://www.ibr.tcu.edu/pubs/datacoll/Forms/wc-Srf-i-sg.PDF

Petranik, S. (Ed.). (2011). Controlling hardcore criminals. In *Hawaii 2011: Our quality of life* (p. 24). Honolulu, HI: Hawaii Business Magazine.

Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology,* 1(4), 435–450.

Pew Charitable Trusts. (2012). *Public opinion on sentencing and corrections policy in America*. Retrieved from http://www.pewtrusts.org/~/media/assets/2012/03/30/pew_nationalsurvey researchpaper_final.pdf.

Project Sentry. (2004). *Who we are and what we do.* Retrieved from http://www.ingham.org/co/ABOUTPROJECTSENTRY.pdf

Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological and training challenges*. Administration and Policy in Mental Health,* 36, 24–34.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Redding, R. E., & Fuller, E. J. (2005). What do juvenile offenders know about being tried as adults? Implications for deterrence. *Juvenile and Family Court Journal,* 55(3), 35–44. Available from: http://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1029&context=wps

Rhoades, B. J., Bumbarger, B. K., & Moore, J. E. (2012). The role of a state-level prevention support system in promoting high-quality implementation and sustainability of evidence-based programs. *American Journal of Community Psychology,* 50, 386–401.

Roberts, J. V. (2013). Public opinion and the nature of community penalties: International findings. *Changing attitudes to punishment: Public opinion, crime and justice.* Roberts, Julian V. & Mike Hough. (eds). Routledge New York: NY.

Rossi, P. H., Lipsey, M., & Freeman, H. (2003). *Evaluation: A systematic approach*. 7th Edition. Newbury Park, CA: Sage.

Sampson, R. J., & Laub, J. H. (1995). *Crime in the Making: Pathways and Turning Points through Life.* Boston, MA: Harvard University Press.

Sampson, R. J., & Bartusch, D. (1998). Legal cynicism and (subcultural?) tolerance of deviance: The neighborhood context of racial differences. *Law & Society Review*, 32, 4, 777–804.

SAS Institute Inc. (2013). *SAS/STAT® 13.1 User's Guide.* Cary, NC: SAS Institute Inc.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Ant in Med,* 152(11), 1–8.

Shepard, D. S., Beaston-Blaakman, A., & Horgan, C. (2003). The ADSS cost study: Costs of substance abuse treatment in the specialty sector. DHHS Publication No. SMA 03-3762, Analytic Series: A-20. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Simourd, D. (1997). Criminal sentiments scale-modified and pride in delinquency scale: Psychometric Properties and construct validity of two measures of criminal attitudes. *Criminal Justice and Behavior*, 24, 1, 52–70.

Snell, C. (2007.) *Fort Bend County Community Supervision and Corrections Special Sanctions Court Program.* Houston, TX: University of Houston Downtown. Retrieved from http://s3.amazonaws.com/static.texastribune.org/media/documents/SANCTIONS_COURT_FINAL_REPORT.pdf?preview.

Snijders, T. A. B. (2002), Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. Journal of Social Structure. Available from http://www.cmu.edu/joss/content/articles/volume3/Snijders.pdf.

Stith, S., Pruitt, I., Dees, J., Fronce, M., Green, N., Som, A., & Linkh, D. (2006). Implementing community-based prevention programming: A review of the literature. *The Journal of Primary Prevention,* 27(6), 599–617.

Tucker, J. A., & Roth, D. L. (2006). Extending the evidence hierarchy to enhance evidence-based practice for substance use disorders. *Addiction*, 101, 918–932.

Tueller, S. J., Van Dorn, R. A., & Bobashev, G. V. (2016). *Visualization of categorical longitudinal and times series data (Report No. MR-0033-1602).* Research Triangle Park, NC: RTI Press.

Tueller, S. J. (2016). longCatEDA: Package for Plotting Categorical Longitudinal and Time-Series Data. R package version 0.21. https://cran.r-project.org/web/packages/longCatEDA/index.html

Van Voorhis, P., & Salisbury, E. J. (2013). *Correctional Counseling and Rehabilitation* (8th ed.). Cincinnati, OH: Anderson Publishing.

Washington State Institute for Public Policy. (2002). *Washington State's implementation of functional family therapy for juvenile offenders: Preliminary findings.* Olympia: Author.

Witte, A. (1980). Estimating the economic model of crime with individual data. *Quarterly Journal of Economics,* 94, 57–84.

Wright, B. R. E., Caspi, A., Moffitt, T. E., & Paternoster, R. (2004). Does the perceived risk of punishment deter criminally prone individuals? Rational choice, self-control, and crime. *Journal of Research in Crime and Delinquency*, 41(2), 180–213.

Yáñez-Correa, A. & Totman, M. (2010). Costly Confinement & Sensible Solutions: Jail Overcrowding in Texas. Retrieved September 6, 2016, from http://www.texascjc.org/sites/default/files/publications/Costly%20Confinement%20Sensible%20Solutions%20Report%20(Oct%202010).pdf.

Zajac, G., Lattimore, P. K., Dawes, D., & Winger, L. (2015). All implementation is local: Initial findings from the process evaluation of the Honest Opportunity Probation with Enforcement (HOPE) demonstration field experiment. *Federal Probation,* 79(1), 31–36.

Appendices